and let v be a vector in $\mathbf{R}^n$. If the limit

$$\lim_{t\to 0; t>0, x_0+tv\in E} \frac{f(x_0+tv)-f(x_0)}{t}$$

exists, we say that f is *differentiable in the direction v at x_0*, and we denote the above limit by $D_v f(x_0)$:

$$D_v f(x_0) := \lim_{t\to 0; t>0} \frac{f(x_0+tv)-f(x_0)}{t}.$$

Remark 17.3.2. One should compare this definition with Definition 17.2.2. Note that we are dividing by a scalar t, rather than a vector, so this definition makes sense, and $D_v f(x_0)$ will be a vector in $\mathbf{R}^m$. It is sometimes possible to also define directional derivatives on the boundary of E, if the vector v is pointing in an "inward" direction (this generalizes the notion of left derivatives and right derivatives from single variable calculus); but we will not pursue these matters here.

Example 17.3.3. If $f : \mathbf{R} \to \mathbf{R}$ is a function, then $D_{+1} f(x)$ is the same as the right derivative of $f(x)$ (if it exists), and similarly $D_{-1} f(x)$ is the same as the left derivative of $f(x)$ (if it exists).

Example 17.3.4. We use the function $f : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $f(x, y) := (x^2, y^2)$ from before, and let $x_0 := (1, 2)$ and $v := (3, 4)$. Then

$$D_v f(x_0) = \lim_{t\to 0; t>0} \frac{f(1+3t, 2+4t) - f(1,2)}{t}$$

$$= \lim_{t\to 0; t>0} \frac{(1+6t+9t^2, 4+16t+16t^2) - (1,4)}{t}$$

$$= \lim_{t\to 0; t>0} (6+9t, 16+16t) = (6, 16).$$

Directional derivatives are connected with total derivatives as follows:

Lemma 17.3.5. *Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, x_0 be an interior point of E, and let v be a vector in*

$\mathbf{R}^n$. *If f is differentiable at x_0, then f is also differentiable in the direction v at x_0, and*

$$D_v f(x_0) = f'(x_0)v.$$

Proof. See Exercise 17.3.1. □

Remark 17.3.6. One consequence of this lemma is that total differentiability implies directional differentiability. However, the converse is not true; see Exercise 17.3.3.

Closely related to the concept of directional derivative is that of *partial derivative*:

Definition 17.3.7 (Partial derivative). Let E be a subset of $\mathbf{R}^n$, let $f : E \to \mathbf{R}^m$ be a function, let x_0 be an interior point of E, and let $1 \leq j \leq m$. Then the *partial derivative of f with respect to the x_j variable* at x_0, denoted $\frac{\partial f}{\partial x_j}(x_0)$, is defined by

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t \to 0; t \neq 0, x_0 + tv \in E} \frac{f(x_0 + te_j) - f(x_0)}{t} = \frac{d}{dt} f(x_0 + te_j)|_{t=0}$$

provided of course that the limit exists. (If the limit does not exist, we leave $\frac{\partial f}{\partial x_j}(x_0)$ undefined).

Informally, the partial derivative can be obtained by holding all the variables other than x_j fixed, and then applying the single-variable calculus derivative in the x_j variable. Note that if f takes values in $\mathbf{R}^m$, then so will $\frac{\partial f}{\partial x_j}$. Indeed, if we write f in components as $f = (f_1, \ldots, f_m)$, it is easy to see (why?) that

$$\frac{\partial f}{\partial x_j}(x_0) = (\frac{\partial f_1}{\partial x_j}(x_0), \ldots, \frac{\partial f_m}{\partial x_j}(x_0)),$$

i.e., to differentiate a vector-valued function one just has to differentiate each of the components separately.

We sometimes replace the variables x_j in $\frac{\partial f}{\partial x_j}$ with other symbols. For instance, if we are dealing with the function $f(x, y) = (x^2, y^2)$, then we might refer to $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ instead of $\frac{\partial f}{\partial x_1}$ and

$\frac{\partial f}{\partial x_2}$. (In this case, $\frac{\partial f}{\partial x}(x, y) = (2x, 0)$ and $\frac{\partial f}{\partial y}(x, y) = (0, 2y)$). One should caution however that one should only relabel the variables if it is absolutely clear which symbol refers to the first variable, which symbol refers to the second variable, etc.; otherwise one may become unintentionally confused. For instance, in the above example, the expression $\frac{\partial f}{\partial x}(x, x)$ is just $(2x, 0)$, however one may mistakenly compute

$$\frac{\partial f}{\partial x}(x, x) = \frac{\partial}{\partial x}(x^2, x^2) = (2x, 2x);$$

the problem here is that the symbol x is being used for more than just the first variable of f. (On the other hand, it is true that $\frac{d}{dx}f(x, x)$ is equal to $(2x, 2x)$; thus the operation of total differentiation $\frac{d}{dx}$ is not the same as that of partial differentiation $\frac{\partial}{\partial x}$).

From Lemma 17.3.5, we know that if a function is differentiable at a point x_0, then all the partial derivatives $\frac{\partial f}{\partial x_j}$ exist at x_0, and that

$$\frac{\partial f}{\partial x_j}(x_0) = f'(x_0)e_j.$$

Also, if $v = (v_1, \ldots, v_n) = \sum_j v_j e_j$, then we have

$$D_v f(x_0) = f'(x_0) \sum_j v_j e_j = \sum_j v_j f'(x_0) e_j$$

(since $f'(x_0)$ is linear) and thus

$$D_v f(x_0) = \sum_j v_j \frac{\partial f}{\partial x_j}(x_0).$$

Thus one can write directional derivatives in terms of partial derivatives, *provided that* the function is actually differentiable at that point.

Just because the partial derivatives exist at a point x_0, we cannot conclude that the function is differentiable there (Exercise 17.3.3). However, if we know that the partial derivatives not only exist, but are continuous, then we can in fact conclude differentiability, thanks to the following handy theorem:

Theorem 17.3.8. *Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, F be a subset of E, and x_0 be an interior point of F. If all the partial derivatives $\frac{\partial f}{\partial x_j}$ exist on F and are continuous at x_0, then f is differentiable at x_0, and the linear transformation $f'(x_0) : \mathbf{R}^n \to \mathbf{R}^m$ is defined by*

$$f'(x_0)(v_j)_{1 \leq j \leq n} = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

Proof. Let $L : \mathbf{R}^n \to \mathbf{R}^m$ be the linear transformation

$$L(v_j)_{1 \leq j \leq m} := \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

We have to prove that

$$\lim_{x \to x_0; x \in E - \{x_0\}} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Let $\varepsilon > 0$. It will suffice to find a radius $\delta > 0$ such that

$$\frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} \leq \varepsilon$$

for all $x \in B(x_0, \delta) \backslash \{x_0\}$. Equivalently, we wish to show that

$$\|f(x) - f(x_0) - L(x - x_0)\| \leq \varepsilon \|x - x_0\|$$

for all $x \in B(x_0, \delta) \backslash \{x_0\}$.

Because x_0 is an interior point of F, there exists a ball $B(x_0, r)$ which is contained inside F. Because each partial derivative $\frac{\partial f}{\partial x_j}$ is continuous on F, there thus exists an $0 < \delta_j < r$ such that $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon/nm$ for every $x \in B(x_0, \delta_j)$. If we take $\delta = \min(\delta_1, \ldots, \delta_n)$, then we thus have $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon/nm$ for every $x \in B(x_0, \delta)$ and every $1 \leq j \leq n$.

Let $x \in B(x_0, \delta)$. We write $x = x_0 + v_1 e_1 + v_2 e_2 + \ldots + v_n e_n$ for some scalars $v_1, \ldots, v_n$. Note that

$$\|x - x_0\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

and in particular we have $|v_j| \leq \|x - x_0\|$ for all $1 \leq j \leq n$. Our task is to show that

$$\|f(x_0 + v_1e_1 + \ldots + v_ne_n) - f(x_0) - \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon\|x - x_0\|.$$

Write f in components as $f = (f_1, f_2, \ldots, f_m)$ (so each f_i is a function from E to $\mathbf{R}$). From the mean value theorem in the x_1 variable, we see that

$$f_i(x_0 + v_1e_1) - f_i(x_0) = \frac{\partial f_i}{\partial x_1}(x_0 + t_ie_1)v_1$$

for some t_i between 0 and v_1. But we have

$$|\frac{\partial f_i}{\partial x_j}(x_0 + t_ie_1) - \frac{\partial f_i}{\partial x_j}(x_0)| \leq \|\frac{\partial f}{\partial x_j}(x_0 + t_ie_1) - \frac{\partial f}{\partial x_j}(x_0)\| \leq \varepsilon/nm$$

and hence

$$|f_i(x_0 + v_1e_1) - f_i(x_0) - \frac{\partial f_i}{\partial x_1}(x_0)v_1| \leq \varepsilon|v_1|/nm.$$

Summing this over all $1 \leq i \leq m$ (and noting that $\|(y_1, \ldots, y_m)\| \leq |y_1| + \ldots + |y_m|$ from the triangle inequality) we obtain

$$\|f(x_0 + v_1e_1) - f(x_0) - \frac{\partial f}{\partial x_1}(x_0)v_1\| \leq \varepsilon|v_1|/n;$$

since $|v_1| \leq \|x - x_0\|$, we thus have

$$\|f(x_0 + v_1e_1) - f(x_0) - \frac{\partial f}{\partial x_1}(x_0)v_1\| \leq \varepsilon\|x - x_0\|/n.$$

A similar argument gives

$$\|f(x_0 + v_1e_1 + v_2e_2) - f(x_0 + v_1e_1) - \frac{\partial f}{\partial x_2}(x_0)v_2\| \leq \varepsilon\|x - x_0\|/n.$$

and so forth up to

$$\|f(x_0 + v_1e_1 + \ldots + v_ne_n) - f(x_0 + v_1e_1 + \ldots + v_{n-1}e_{n-1}) \\ - \frac{\partial f}{\partial x_n}(x_0)v_n\| \leq \varepsilon\|x - x_0\|/n.$$

If we sum these n inequalities and use the triangle inequality $\|x+y\| \leq \|x\|+\|y\|$, we obtain a telescoping series which simplifies to

$$\|f(x_0+v_1e_1+\ldots+v_ne_n)-f(x_0)-\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(x_0)v_j\| \leq \varepsilon\|x-x_0\|$$

as desired. □

From Theorem 17.3.8 and Lemma 17.3.5 we see that if the partial derivatives of a function $f : E \to \mathbf{R}^m$ exist and are continuous on some set F, then all the directional derivatives also exist at every interior point x_0 of F, and we have the formula

$$D_{(v_1,\ldots,v_n)}f(x_0) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

In particular, if $f : E \to \mathbf{R}$ is a real-valued function, and we define the *gradient* $\nabla f(x_0)$ of f at x_0 to be the n-dimensional row vector $\nabla f(x_0) := (\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0))$, then we have the familiar formula

$$D_v f(x_0) = v \cdot \nabla f(x_0)$$

whenever x_0 is in the interior of the region where the gradient exists and is continuous.

More generally, if $f : E \to \mathbf{R}^m$ is a function taking values in $\mathbf{R}^m$, with $f = (f_1, \ldots, f_m)$, and x_0 is in the interior of the region where the partial derivatives of f exist and are continuous, then we have from Theorem 17.3.8 that

$$\begin{aligned} f'(x_0)(v_j)_{1\leq j\leq n} &= \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0) \\ &= (\sum_{j=1}^{n} v_j \frac{\partial f_i}{\partial x_j}(x_0))_{i=1}^{m}, \end{aligned}$$

which we can rewrite as

$$L_{Df(x_0)}(v_j)_{1\leq j\leq n}$$

where $Df(x_0)$ is the $m \times n$ matrix

$$Df(x_0) := (\frac{\partial f_i}{\partial x_j}(x_0))_{1 \leq i \leq m; 1 \leq j \leq n}$$
$$= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \dots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \dots & \frac{\partial f_2}{\partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \dots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}.$$

Thus we have

$$(D_v f(x_0))^T = (f'(x_0)v)^T = Df(x_0)v^T.$$

The matrix $Df(x_0)$ is sometimes also called the *derivative matrix* or *differential matrix* of f at x_0, and is closely related to the total derivative $f'(x_0)$. One can also write Df as

$$Df(x_0) = (\frac{\partial f}{\partial x_1}(x_0)^T, \frac{\partial f}{\partial x_2}(x_0)^T, \dots, \frac{\partial f}{\partial x_n}(x_0)^T),$$

i.e., each of the columns of $Df(x_0)$ is one of the partial derivatives of f, expressed as a column vector. Or one could write

$$Df(x_0) = \begin{pmatrix} \nabla f_1(x_0) \\ \nabla f_2(x_0) \\ \vdots \\ \nabla f_m(x_0) \end{pmatrix}$$

i.e., the rows of $Df(x_0)$ are the gradient of various components of f. In particular, if f is scalar-valued (i.e., $m = 1$), then Df is the same as ∇f.

Example 17.3.9. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the function $f(x, y) = (x^2 + xy, y^2)$. Then $\frac{\partial f}{\partial x} = (2x + y, 0)$ and $\frac{\partial f}{\partial y} = (x, 2y)$. Since these partial derivatives are continuous on $\mathbf{R}^2$, we see that f is differentiable on all of $\mathbf{R}^2$, and

$$Df(x, y) = \begin{pmatrix} 2x + y & x \\ 0 & 2y \end{pmatrix}.$$

Thus for instance, the directional derivative in the direction (v, w) is

$$D_{(v,w)}f(x, y) = ((2x + y)v + xw, 2yw).$$

Exercise 17.3.1. Prove Lemma 17.3.5. (This will be similar to Exercise 17.1.3).

Exercise 17.3.2. Let E be a subset of $\mathbf{R}^n$, let $f : E \to \mathbf{R}^m$ be a function, let x_0 be an interior point of E, and let $1 \leq j \leq n$. Show that $\frac{\partial f}{\partial x_j}(x_0)$ exists if and only if $D_{e_j}f(x_0)$ and $D_{-e_j}f(x_0)$ exist and are negatives of each other (thus $D_{e_j}f(x_0) = D_{-e_j}f(x_0)$); furthermore, one has $\frac{\partial f}{\partial x_j}(x_0) = D_{e_j}f(x_0)$ in this case.

Exercise 17.3.3. Let $f : \mathbf{R}^2 \to \mathbf{R}$ be the function defined by $f(x, y) := \frac{x^3}{x^2+y^2}$ when $(x, y) \neq (0, 0)$, and $f(0, 0) := 0$. Show that f is not differentiable at $(0, 0)$, despite being differentiable in every direction $v \in \mathbf{R}^2$ at $(0, 0)$. Explain why this does not contradict Theorem 17.3.8.

Exercise 17.3.4. Let $f : \mathbf{R}^n \to \mathbf{R}^m$ be a differentiable function such that $f'(x) = 0$ for all $x \in \mathbf{R}^n$. Show that f is constant. (Hint: you may use the mean-value theorem or fundamental theorem of calculus for one-dimensional functions, but bear in mind that there is no direct analogue of these theorems for several-variable functions. I would not advise proceeding via first principles.) For a tougher challenge, replace the domain $\mathbf{R}^n$ by an open connected subset Ω of $\mathbf{R}^n$.

17.4 The several variable calculus chain rule

We are now ready to state the several variable calculus chain rule. Recall that if $f : X \to Y$ and $g : Y \to Z$ are two functions, then the composition $g \circ f : X \to Z$ is defined by $g \circ f(x) := g(f(x))$ for all $x \in X$.

Theorem 17.4.1 (Several variable calculus chain rule). *Let E be a subset of $\mathbf{R}^n$, and let F be a subset of $\mathbf{R}^m$. Let $f : E \to F$ be a function, and let $g : F \to \mathbf{R}^p$ be another function. Let x_0 be a point in the interior of E. Suppose that f is differentiable at x_0, and that $f(x_0)$ is in the interior of F. Suppose also that g is* *differentiable at $f(x_0)$. Then $g \circ f : E \to \mathbf{R}^p$ is also differentiable at x_0, and we have the formula*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

Proof. See Exercise 17.4.3. □

One should compare this theorem with the single-variable chain rule, Theorem 10.1.15; indeed one can easily deduce the single-variable rule as a consequence of the several-variable rule.

Intuitively, one can think of the several variable chain rule as follows. Let x be close to x_0. Then Newton's approximation asserts that

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0)$$

and in particular $f(x)$ is close to $f(x_0)$. Since g is differentiable at $f(x_0)$, we see from Newton's approximation again that

$$g(f(x)) - g(f(x_0)) \approx g'(f(x_0))(f(x) - f(x_0)).$$

Combining the two, we obtain

$$g \circ f(x) - g \circ f(x_0) \approx g'(f(x_0))f'(x_0)(x - x_0)$$

which then should give $(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$. This argument however is rather imprecise; to make it more precise one needs to manipulate limits rigourously; see Exercise 17.4.3.

As a corollary of the chain rule and Lemma 17.1.16 (and Lemma 17.1.13), we see that

$$D(g \circ f)(x_0) = Dg(f(x_0))Df(x_0);$$

i.e., we can write the chain rule in terms of matrices and matrix multiplication, instead of in terms of linear transformations and composition.

Example 17.4.2. Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$ be differentiable functions. We form the combined function $h : \mathbf{R}^n \to \mathbf{R}^2$

by defining $h(x) := (f(x), g(x))$. Now let $k : \mathbf{R}^2 \to \mathbf{R}$ be the multiplication function $k(a, b) := ab$. Note that

$$Dh(x_0) = \begin{pmatrix} \nabla f(x_0) \\ \nabla g(x_0) \end{pmatrix}$$

while

$$Dk(a, b) = (b, a)$$

(why?). By the chain rule, we thus see that

$$D(k \circ h)(x_0) = (g(x_0), f(x_0)) \begin{pmatrix} \nabla f(x_0) \\ \nabla g(x_0) \end{pmatrix} = g(x_0)\nabla f(x_0) + f(x_0)\nabla g(x_0)$$

But $k \circ h = fg$ (why?), and $D(fg) = \nabla(fg)$. We have thus proven the *product rule*

$$\nabla(fg) = g\nabla f + f\nabla g.$$

A similar argument gives the sum rule $\nabla(f + g) = \nabla f + \nabla g$, or the difference rule $\nabla(f - g) = \nabla f - \nabla g$, as well as the quotient rule (Exercise 17.4.4). As you can see, the several variable chain rule is quite powerful, and can be used to deduce many other rules of differentiation.

We do record one further useful application of the chain rule. Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. From Exercise 17.4.1 we observe that T is continuously differentiable at every point, and in fact $T'(x) = T$ for every x. (This equation may look a little strange, but perhaps it is easier to swallow if you view it in the form $\frac{d}{dx}(Tx) = T$). Thus, for any differentiable function $f : E \to \mathbf{R}^n$, we see that $Tf : E \to \mathbf{R}^m$ is also differentiable, and hence by the chain rule

$$(Tf)'(x_0) = T(f'(x_0)).$$

This is a generalization of the single-variable calculus rule $(cf)' = c(f')$ for constant scalars c.

Another special case of the chain rule which is quite useful is the following: if $f : \mathbf{R}^n \to \mathbf{R}^m$ is some differentiable function, and

$x_j : \mathbf{R} \to \mathbf{R}$ are differentiable functions for each $j = 1, \ldots n$, then

$$\frac{d}{dt} f(x_1(t), x_2(t), \ldots, x_n(t)) = \sum_{j=1}^{n} x_j'(t) \frac{\partial f}{\partial x_j}(x_1(t), x_2(t), \ldots, x_n(t)).$$

(Why is this a special case of the chain rule?).

Exercise 17.4.1. Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Show that T is continuously differentiable at every point, and in fact $T'(x) = T$ for every x. What is DT?

Exercise 17.4.2. Let E be a subset of $\mathbf{R}^n$. Prove that if a function $f : E \to \mathbf{R}^m$ is differentiable at an interior point x_0 of E, then it is also continuous at x_0. (Hint: use Exercise 17.1.4.)

Exercise 17.4.3. Prove Theorem 17.4.1. (Hint: you may wish to review the proof of the ordinary chain rule in single variable calculus, Theorem 10.1.15. The easiest way to proceed is by using the sequence-based definition of limit (see Proposition 14.1.5(b)), and use Exercise 17.1.4.)

Exercise 17.4.4. State and prove some version of the quotient rule for functions of several variables (i.e., functions of the form $f : E \to \mathbf{R}$ for some subset E of $\mathbf{R}^n$). In other words, state a rule which gives a formula for the gradient of f/g; compare your answer with Theorem 10.1.13(h). Be sure to make clear what all your assumptions are.

Exercise 17.4.5. Let $\vec{x} : \mathbf{R} \to \mathbf{R}^3$ be a differentiable function, and let $r : \mathbf{R} \to \mathbf{R}$ be the function $r(t) := \|\vec{x}(t)\|$, where $\|\vec{x}\|$ denotes the length of $\vec{x}$ as measured in the usual l^2 metric. Let t_0 be a real number. Show that if $r(t_0) \neq 0$, then r is differentiable at t_0, and

$$r'(t_0) = \frac{\vec{x}'(t_0) \cdot \vec{x}(t_0)}{r(t_0)}.$$

(Hint: use Theorem 17.4.1.)

17.5 Double derivatives and Clairaut's theorem

We now investigate what happens if one differentiates a function twice.

Definition 17.5.1 (Twice continuous differentiability). Let E be an open subset of $\mathbf{R}^n$, and let $f : E \to \mathbf{R}^m$ be a function. We say that f is *continuously differentiable* if the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ exist and are continuous on E. We say that f is *twice continuously differentiable* if it is continuously differentiable, and the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are themselves continuously differentiable.

Remark 17.5.2. Continuously differentiable functions are sometimes called C^1 functions; twice continuously differentiable functions are sometimes called C^2 functions. One can also define C^3, C^4, etc. but we shall not do so here.

Example 17.5.3. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the function $f(x, y) = (x^2 + xy, y^2)$. Then f is continuously differentiable because the partial derivatives $\frac{\partial f}{\partial x}(x, y) = (2x + y, 0)$ and $\frac{\partial f}{\partial y}(x, y) = (x, 2y)$ exist and are continuous on all of $\mathbf{R}^2$. It is also twice continuously differentiable, because the double partial derivatives $\frac{\partial}{\partial x}\frac{\partial f}{\partial x}(x, y) = (2, 0)$, $\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(x, y) = (1, 0)$, $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x, y) = (1, 0)$, $\frac{\partial}{\partial y}\frac{\partial f}{\partial y}(x, y) = (0, 2)$ all exist and are continuous.

Observe in the above example that the double derivatives $\frac{\partial}{\partial y}\frac{\partial f}{\partial x}$ and $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}$ are the same. This is a in fact a general phenomenon:

Theorem 17.5.4 (Clairaut's theorem). *Let E be an open subset of $\mathbf{R}^n$, and let $f : E \to \mathbf{R}$ be a twice continuously differentiable function on E. Then we have $\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x_0) = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x_0)$ for all $1 \leq i, j \leq n$.*

Proof. The claim is trivial if $i = j$, so we shall assume that $i \neq j$. We shall prove the theorem for $x_0 = 0$; the general case is similar. (Actually, once one proves Clairaut's theorem for $x_0 = 0$, one can immediately obtain it for general x_0 by applying the theorem with $f(x)$ replaced by $f(x - x_0)$.)

Let a be the number $a := \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(0)$, and a' denote the quantity $a' := \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(0)$. Our task is to show that $a' = a$.

Let $\varepsilon > 0$. Because the double derivatives of f are continuous, we can find a $\delta > 0$ such that

$$|\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x) - a| \leq \varepsilon$$

and

$$|\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) - a'| \leq \varepsilon$$

whenever $|x| \leq 2\delta$.

Now we consider the quantity

$$X := f(\delta e_i + \delta e_j) - f(\delta e_i) - f(\delta e_j) - f(0).$$

From the fundamental theorem of calculus in the e_i variable, we have

$$f(\delta e_i + \delta e_j) - f(\delta e_j) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j)\, dx_i$$

and

$$f(\delta e_i) - f(0) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i)\, dx_i$$

and hence

$$X \doteq \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i)\, dx_i.$$

But by the mean value theorem, for each x_i we have

$$\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) = \delta \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x_i e_i + x_j e_j)$$

for some $0 \leq x_j \leq \delta$. By our construction of δ, we thus have

$$|\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a| \leq \varepsilon\delta.$$

Integrating this from 0 to δ, we thus obtain

$$|X - \delta^2 a| \leq \varepsilon\delta^2.$$

We can run the same argument with the rôle of i and j reversed (note that X is symmetric in i and j), to obtain

$$|X - \delta^2 a'| \leq \varepsilon\delta^2.$$

From the triangle inequality we thus obtain

$$|\delta^2 a - \delta^2 a'| \leq 2\varepsilon\delta^2,$$

and thus

$$|a - a'| \leq 2\varepsilon.$$

But this is true for all $\varepsilon > 0$, and a and a' do not depend on ε, and so we must have $a = a'$, as desired. □

One should caution that Clairaut's theorem fails if we do not assume the double derivatives to be continuous; see Exercise 17.5.1.

Exercise 17.5.1. Let $f : \mathbf{R}^2 \to \mathbf{R}$ be the function defined by $f(x, y) := \frac{xy^3}{x^2+y^2}$ when $(x, y) \neq (0, 0)$, and $f(0, 0) := 0$. Show that f is continuously differentiable, and the double derivatives $\frac{\partial}{\partial y}\frac{\partial f}{\partial x}$ and $\frac{\partial}{\partial x}\frac{\partial f}{\partial y}$ exist, but are not equal to each other at $(0, 0)$. Explain why this does not contradict Clairaut's theorem.

17.6 The contraction mapping theorem

Before we turn to the next topic - namely, the inverse function theorem - we need to develop a useful fact from the theory of complete metric spaces, namely the contraction mapping theorem.

Definition 17.6.1 (Contraction). Let (X, d) be a metric space, and let $f : X \to X$ be a map. We say that f is a *contraction* if we have $d(f(x), f(y)) \leq d(x, y)$ for all $x, y \in X$. We say that f is a *strict contraction* if there exists a constant $0 < c < 1$ such that $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$; we call c the *contraction constant* of f.

Examples 17.6.2. The map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) := x+1$ is a contraction but not a strict contraction. The map $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) := x/2$ is a strict contraction. The map $f : [0, 1] \to [0, 1]$ defined by $f(x) := x - x^2$ is a contraction but not a strict contraction. (For justifications of these statements, see Exercise 17.6.5.