Unit 3: Differentiability and the Gradient

3.3.1(L)

In our study of the calculus of a single real variable, we saw that if f'(a) existed then f was continuous at x = a. In other words, we saw that the property of being differentiable implied the property of continuity. (In geometric terms, we are saying that if a curve is smooth then it must be unbroken.)

Analogously, one might expect that, in two variables, if $f_x(a,b)$ and $f_y(a,b)$ exist then f must be continuous at the point (a,b). The aim of this exercise is to show that this need not be true. Before we actually do the exercise, let us observe that, in the case of one independent variable, the fact that f'(a) existed guaranteed that the directional derivative existed in each direction, since in this case there are only two directions. In the case of two independent variables, however, the fact that the directional derivatives exist in both the horizontal and vertical directions tells us nothing about what is happening in other directions.

Rather than continue with an abstract discussion, we shall show in this exercise that the function g has the property that $g_x(0,0)$ and $g_y(0,0)$ both exist - yet g is <u>not</u> continuous at the point (0,0). This, in turn, is enough to show that the assumption that g is continuous at a point, after we know that both g_x and g_y exist at that point, is not redundant.

As far as the solution of our exercise is concerned, part of it is trivial, since we have already proven that g was not continuous at (0,0) in Exercise 3.1.8(L).

Thus, we need only show that $g_x(0,0)$ and $g_y(0,0)$ exist. The easiest approach is to work directly from the definition of g_x and g_y.

For example $g_x(a,b)$ means $\lim\limits_{\Delta x \to 0} \dfrac{g(a+\Delta x,b)-g(a,b)}{\Delta x}$ so that we have

3.3.1 continued

$$g_x(0,0) = \lim_{\Delta x \to 0} \left[\frac{g(\Delta x,0)-g(0,0)}{\Delta x} \right]$$

$$= \lim_{\Delta x \to 0} \left[\frac{\frac{2\Delta x(0)}{\Delta x^2+0^2} - 0}{\Delta x} \right]$$

$$= \lim_{\Delta x \to 0} \left[\frac{0}{\Delta x} \right]$$

$$= 0^*$$

Since g is symmetric in x and y [i.e., $g(x,y)=g(y,x)$], it follows at once that

$g_y(0,0) = 0$ also.

(If the symmetry property doesn't strike your fancy, mimic our procedure for showing $g_x(0,0) = 0$ to prove $g_y(0,0) = 0$.)

This exercise sets the stage for a deeper analysis of what happens when we deal with functions of several variables. Without going into detail about why it is so (the proof of Theorem 1 in Section 15.4 of the text shows this as does our optional exercise 3.3.9), the fact remains that if $f_x(a,b)$ and $f_y(a,b)$ exist and if both f_x and f_y are continuous at (a,b), then f will be continuous at (a,b) (although in the text this conclusion is stated as part of the hypotheses). From a somewhat intuitive point of view, the fact

---

*Do not be tempted to say that since $g(0,0)=0$, $\frac{\partial}{\partial x} g(x,y)\Big|_{(0,0)} = \frac{\partial 0}{\partial x} = 0$. This would be true if $g(x,y)=0$. In the 1-dimensional case, notice that if $g(x) = x^2-2x$ then $f'(x)=2x-2$. Hence, $g'(0)=-2$ even though $g(0)=0$. (Pictorially, $g(x)=x^2-2x$ says that the curve $y=g(x)$ passes through (0,0). It does not say that the slope of the curve at this point is 0). Perhaps a better example might be the one we used in Part 1 of the course where f was defined by

$$f(x) = \begin{cases} \frac{x^2-9}{x-3} & \text{if } x \neq 3 \\ 6 & \text{if } x = 3 \end{cases}$$

. This means that $f(x) \equiv x+3$. Notice that while $f(3) = 6$, $f'(3) = 1$ not 0.

3.3.1 continued

that f_x and f_y are continuous at (a,b) means that the function is "well behaved" in a neighborhood of (a,b).

To go one step further, and this will be discussed in somewhat more detail in this and the next unit, it turns out that if we want to think of f as being differentiable, it is not enough to think only of f_x and f_y existing; they must also be continuous. Quite in general, if $f(x_1,\ldots x_n)$ has the property that $f_{x_1},\ldots,$ and f_{x_n} exist and are continuous at $(a_1,\ldots,a_n)$ then we say that f is differentiable at $(a_1,\ldots,a_n)$. In this event, f is continuous at $(a_1,\ldots,a_n)$. f need not be continuous if $f_{x_1},\ldots,$ and f_{x_n} are not all continuous at $(a_1,\ldots,a_n)$.

To try this idea out in terms of our present example, since $g_x(0,0)$ and $g_y(0,0)$ exist but g is not continuous at (0,0), it must be that either g_x or g_y is not continuous at (0,0).

Let us look at $g_x(x,y)$ and see whether $\lim_{(x,y)\to(0,0)} g_x(x,y)$ exists.

Since

$$g(x,y) = \begin{cases} \dfrac{2xy}{x^2+y^2} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases}$$

we obtain

$$g_x(x,y) = \begin{cases} \dfrac{(x^2+y^2)2y-2xy(2x)}{(x^2+y^2)^2} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases}$$

or

3.3.1 continued

$$g_x(x,y) = \begin{cases} \dfrac{2y(y^2-x^2)}{(x^2+y^2)^2} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases} \qquad (1)$$

Now, one way of viewing $\lim\limits_{(x,y)\to(0,0)} g_x(x,y)$ is by first letting $x\to 0$ (holding $y\neq 0$ fixed) and then letting $y\to 0$.

If we do this, then

$$\lim_{(x,y)\to(0,0)} g_x(x,y) = \lim_{y\to 0}\left[\lim_{x\to 0} g_x(x,y)\right],$$

and from (1) this yields

$$\lim_{(x,y)\to(0,0)} g_x(x,y) = \lim_{y\to 0}\left[\lim_{x\to 0}\left\{\frac{2y(y^2-x^2)}{(x^2+y^2)^2}\right\}\right]$$

$$= \lim_{y\to 0}\left[\frac{2y^3}{y^4}\right]$$

$$= \lim_{y\to 0}\left[\frac{2}{y}\right]$$

$$= \pm\infty \text{ (depending on whether } y\to 0^+ \text{ or } y\to 0^-) \qquad (2)$$

3.3.1 continued

Since $g_x(0,0) = 0$, (2) shows us that $\lim_{(x,y)\to(0,0)} g_x(x,y) \neq g_x(0,0)$*, so that by the definition of continuity, g_x is not continuous at (0,0).

3.3.2

a. $w = f(x,y) = x^2+x^3y+y^3$

Therefore

$$f(x+\Delta x,y+\Delta y)=(x+\Delta x)^2+(x+\Delta x)^3(y+\Delta y)+(y+\Delta y)^3$$

$$= x^2+2x\Delta x+\overline{\Delta x}^2+(x^3+3x^2\Delta x+3x\overline{\Delta x}^2+\overline{\Delta x}^3)(y+\Delta y)$$

$$+ (y^3+3y^2\Delta y+3y\overline{\Delta y}^2+\overline{\Delta y}^3)$$

$$= x^2+2x\Delta x+\overline{\Delta x}^2+x^3y+3x^2y\Delta x+3xy\overline{\Delta x}^2+y\overline{\Delta x}^3$$

$$+x^3\Delta y+3x^2\Delta x\Delta y+3x\overline{\Delta x}^2\Delta y+\overline{\Delta x}^3\Delta y$$

$$+ y^3+3y^2\Delta y+3y\overline{\Delta y}^2+\overline{\Delta y}^3$$

therefore

$f(x+\Delta x,y+\Delta y)-f(x,y)$

$$= (2x+3x^2y)\Delta x+(x^3+3y^2)\Delta y+(\Delta x+3xy\Delta x+y\overline{\Delta x}^2+3x^2\Delta y+3x\Delta x\Delta y+\overline{\Delta x}^2\Delta y)\Delta x$$

$$+ (3y\Delta y+\overline{\Delta y}^2)\Delta y$$

*Notice that we picked <u>one special path</u> to compute $\lim_{(x,y)\to(0,0)} g_x(x,y)$. Since along this path, the limit was not equal to $g_x(0,0)$, it guarantees that g_x is not continuous at (0,0), for if it were, then $\lim_{(x,y)\to(0,0)} g_x(x,y)$ would equal $g_x(0,0)$ along <u>every</u> path.

3.3.2 continued

and since

$f_x(x,y) = 2x+3x^2y$ and $f_y(x,y)=x^3+3y^2$,

we see that

$f(x+\Delta x,y+\Delta y)-f(x,y)=f_x(x,y)\Delta x+f_y(x,y)\Delta y +k_1\Delta x+k_2\Delta y$

where

$k_1 = \Delta x+3xy\Delta x+y\overline{\Delta x}^2+3x^2\Delta y+3x\Delta x\Delta y+\overline{\Delta x}^2\Delta y$

$k_2 = 3y\Delta y+\overline{\Delta y}^2$

Hence, $k_1,k_2\to 0$ as $\Delta x_1, \Delta y\to 0$.

b. To estimate $(1.001)^2 + (1.001)^3(1.002) + (1.002)^3$ we may use (a) with $x=y=1$, $\Delta x=0.001$, and $\Delta y=0.002$

$f(1,1) = 3$

Then

$f_x(1,1) = 5$

$f_y(1,1) = 4$

so that

$f_x(1,1)\Delta x + f_y(1,1)\Delta y = 5(0.001) + 4(0.002) = 0.013$

Therefore

$f(1.001, 1.002) \approx f(1,1) + f_x(1,1)\Delta x + f_y(1,1)\Delta y$

$\approx 3 + 0.013 = 3.013$

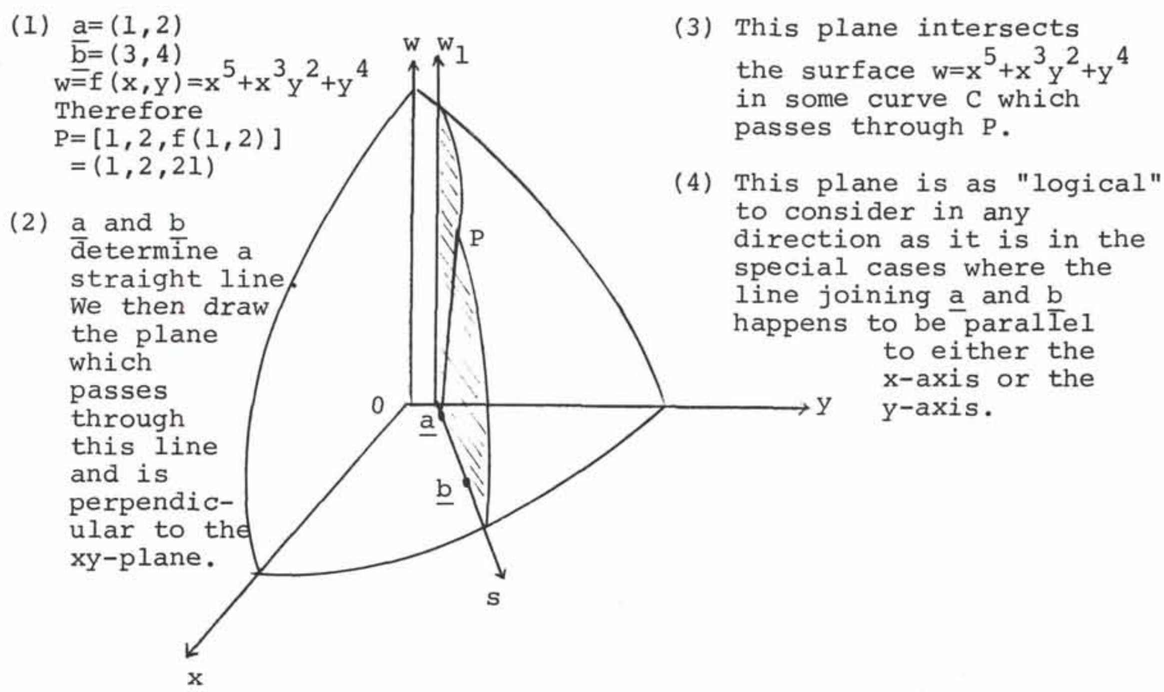
3.3.3(L)

The aim of this exercise is to help make it clear that f_x and f_y are essentially just two special directional derivatives, but that the concept of a directional derivative makes just as good sense regardless of the direction. At the same time, we want to

3.3.3(L) continued

make it clear that if f_x and f_y exist and are continuous at the point (a_1,a_2) then the directional derivative of f at (a_1,a_2) exists in every direction, and, in fact, the directional derivative in each direction can be quickly obtained by knowing the values of $f_x(\underline{a})$ and $f_y(\underline{a})$.

Before continuing further, perhaps a picture or two will be helpful.



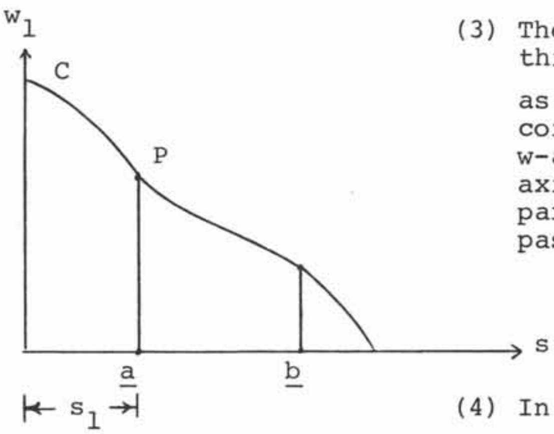
(Figure 1)

a. If we now label by s the direction of the line from $\underline{a}$ to $\underline{b}$, the curve C lies in the w_1s-plane. That is, our shaded portion of Figure 1 may be viewed as

3.3.3(L) continued

(1) The slope of C at P is $\left.\frac{dw_1}{ds}\right|_{s=\underline{a}}$ *

(2) The complication that exists here but didn't when we computed $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$ is that while the x- and y-axes pass through 0, the s-axis need not.



(3) Therefore, we may think of the w_1-axis as being shifted to coincide with the w-axis, and the s-axis as being shifted parallel to itself to pass through 0.

(4) In this way $\frac{dw_1}{ds} = \frac{dw}{ds}$ since the curve C is not affected (only its position in space).

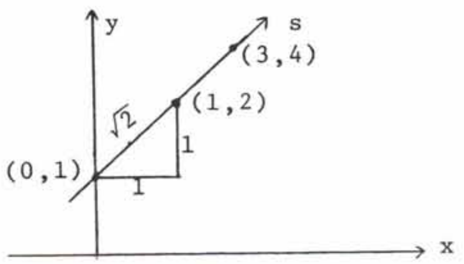
(Figure 2)

The interesting point is that we can compute $\frac{dw}{ds}$ without having to refer to a 3-dimensional diagram. Namely, we know that viewed in the xy-plane, s is the line determined by (1,2) and (3,4); hence, its slope is 1. Its equation, therefore, is $\frac{y-2}{x-1} = 1$, or $y = x+1$.

Thus, while w is a function of the two independent variables x and y, once we restrict our consideration to those points (x,y)

---

*Notice that $\underline{a}=(a_1,a_2)$ refers to coordinates in the xy-plane. In the w_1s-plane $\underline{a}$ would have the form $(s_1,0)$. In our particular example, $\underline{a}=(1,2)\to s_1 = \sqrt{2}$. That is:



3.3.3(L) continued

for which y = x+1, x and y are no longer independent. In other words, in the direction of s, w can be expressed entirely in terms of x (or of y or, for that matter, of s). In particular, our definition of w leads to the fact that in the direction of s,

$$w = x^5+x^3(x+1)^2+(x+1)^4 \qquad (1)$$

whereupon expanding and collection of terms yields

$$w = 2x^5+3x^4+5x^3+6x^2+4x+1 \qquad (2)$$

[Notice that (1) and (2) are special cases of $w = x^5+x^3y^2+y^4$ for a particular path. In general, if s had been any line drawn from (1,2) but not parallel to the y-axis, then the line s would have the form y = mx+b, whereupon (1) would have been

$$w = x^5+x^3(mx+b)^2+(mx+b)^4$$

so that it should be clear that how w varies with x does indeed depend on the direction of the line from (1,2).]

If we differentiate w in (1) with respect to x, (See Remarks in Figure 3) we obtain

$$\frac{dw}{dx} = 10x^4+12x^3+15x^2+12x+4 \qquad (3)$$

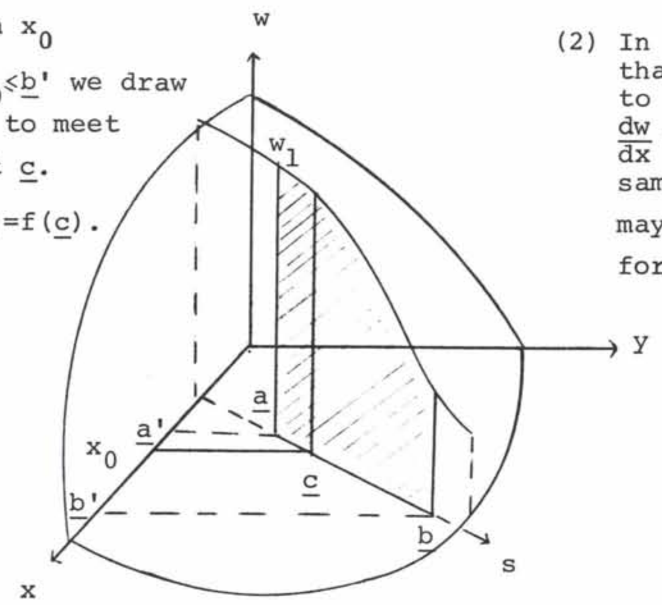
and evaluating (3) for x=1 [i.e., at the point (1,2)] we obtain:

$$\left.\frac{dw}{dx}\right|_{(1,2)} = 53 \qquad (4)$$

3.3.3(L) continued

(1) Given x_0

$\underline{a}' \leq x_0 \leq \underline{b}'$ we draw $x=x_0$ to meet $\underline{ab}$ at $\underline{c}$. Then $w(x_0)=f(\underline{c})$.



(2) In the special case that $\underline{ab}$ is parallel to the x-axis $\frac{dw}{dx}$ happens to be the same as $\frac{\partial w}{\partial x}$, but we may talk about $\frac{dw}{dx}$ for any direction s.

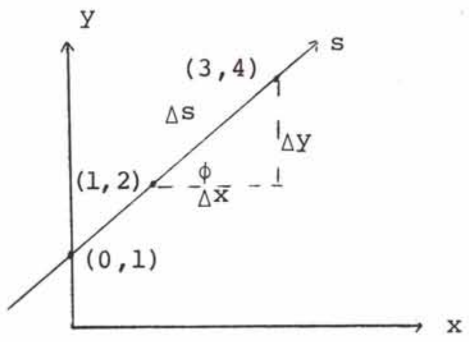
(Figure 3)

The trouble with (4) is that it is the right answer to the wrong problem! We wanted the value of $\frac{dw}{ds}$ not of $\frac{dw}{dx}$. This matter is easily remedied by our use of the chain rule (and here it is important to understand fully that w is a function of the <u>single</u> real variable s; for while s does depend on both x and y, along any fixed line the value of x determines the value of y, so that we have indeed only one degree of freedom). We have

$$\frac{dw}{ds} = \left(\frac{dw}{dx}\right)\left(\frac{dx}{ds}\right) \tag{5}$$

and, pictorially, it is easy to visualize $\frac{dw}{ds}$; namely,

3.3.3(L) continued



$\tan\phi$ = slope of line s=1 or $\phi = \frac{\pi}{4}$

therefore

$$\frac{\Delta x}{\Delta s} = \cos\frac{\pi}{4} = \frac{1}{2}\sqrt{2}$$

(Figure 4)*

Since $\frac{\Delta x}{\Delta s}$ is the <u>constant</u> $\frac{1}{2}\sqrt{2}$, $\frac{dx}{ds} = \lim_{\Delta s \to 0} \frac{\Delta x}{\Delta s} = \frac{1}{2}\sqrt{2}$, and combining this with (4), equation (5) yields

$$\left.\frac{dw}{ds}\right|_{\underline{a}=(1,2)} \quad 53\left(\frac{1}{2}\sqrt{2}\right) = \frac{53\sqrt{2}}{2} \approx 37.5$$

That is, the derivative of w in the direction s which joins (1,2) with (3,4) is $\frac{53\sqrt{2}}{2}$. (With reference to Figure 1, the slope of the curve C at the point P is $\frac{53\sqrt{2}}{2}$.)

b. Since, in our present exercise, w is a polynomial in x and y, it is readily verified that w, w_x, w_y all exist and are continuous at $\underline{a}$. Thus, the fundamental result of this assignment holds; namely,

*Note that Figure 4 is independent of w. That is, all Figure 3 takes into account is what is happening in the xy-plane.

3.3.3(L) continued

$$\Delta w = f_x(1,2)\Delta x + f_y(1,2)\Delta y + k_1\Delta x + k_2\Delta y$$

where k_1 and $k_2 \to 0$ as Δx and $\Delta y \to 0$.

Therefore,

$$\frac{\Delta w}{\Delta s} = f_x(1,2)\ \frac{\Delta x}{\Delta s} + f_y(1,2)\ \frac{\Delta y}{\Delta s} + k_1\ \frac{\Delta x}{\Delta s} + k_2\ \frac{\Delta y}{\Delta s}$$

(This was derived in both the text and the lecture, but it is important enough so that you should see it again.)

Letting $\Delta s \to 0$ makes both Δx and $\Delta y \to 0$, so that k_1 and $k_2 \to 0$

$$\left.\frac{dw}{ds}\right|_{(1,2)} = \lim_{\Delta s \to 0}\ \frac{\Delta w}{\Delta s} = f_x(1,2)\ \frac{dx}{ds} + f_y(1,2)\ \frac{dy}{ds} + 0\ \frac{dx}{ds} + 0\ \frac{dy}{ds}$$

or since $\frac{dx}{ds} = \cos\phi$ and $\frac{dy}{ds} = \sin\phi$,

$$\left.\frac{dw}{ds}\right|_{(1,2)} = f_x(1,2)\ \cos\phi + f_y(1,2)\ \sin\phi \tag{6}$$

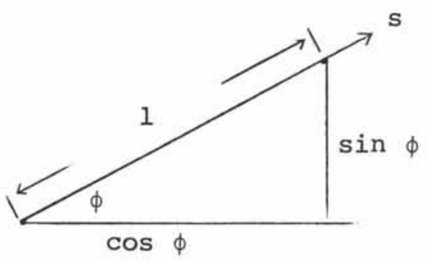
or in dot product notation

$$\left.\frac{dw}{ds}\right|_{(1,2)} = [f_x(1,2), f_y(1,2)] \cdot [\cos\phi,\ \sin\phi] \tag{7}$$

Our first factor on the right side of (7) is $\vec{\nabla} f(1,2)$ which depends only on f and the point (1,2) while our second factor is the unit vector in the direction of s which is independent of f and (1,2).

Pictorially,

3.3.3(L) continued



Thus, (7) becomes

$$\left.\frac{dw}{ds}\right|_{(1,2)} = \vec{\nabla}f(1,2) \cdot \vec{u}_s \qquad (8)$$

Now, since $f(x,y) = x^5+x^3y^2+y^4$, we have $f_x(x,y) = 5x^4+3x^2y^2$ and $f_y(x,y) = 2x^3y+4y^3$. Hence, $f_x(1,2) = 17$ and $f_y(x,y) = 36$. Therefore

$$\vec{\nabla}f(1,2) = (17,36) \quad . \qquad (9)$$

Since the vector from (1,2) to (3,4) is

$$2\vec{1} + 2\vec{j}, \vec{u}_s = \frac{2\vec{1}+2\vec{j}}{\|2\vec{1}+2\vec{j}\|} = \frac{2\vec{1}+2\vec{j}}{\sqrt{2^2+2^2}} = \frac{1}{2}\sqrt{2}\ (\vec{1}+\vec{j}) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$$

Putting this result and the result of (9) into (8) we obtain

$$\left.\frac{dw}{ds}\right|_{(1,2)} = (17,36) \cdot \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$$

$$= \frac{17\sqrt{2}}{2} + \frac{36\sqrt{2}}{2}$$

$$= \frac{53\sqrt{2}}{2} \quad ,$$

3.3.3(L) continued

which agrees with our result in (a).

c. From (8) and (9)

$$\left.\frac{dw}{ds}\right|_{(1,2)} = (17,36) \cdot \vec{u}_s \tag{10}$$

If our direction s is now determined by the vector from (1,2) to (4,6), we have,

$$\vec{u}_s = \frac{3\vec{i}+4\vec{j}}{\|3\vec{i}+4\vec{j}\|} = \frac{3\vec{i}+4\vec{j}}{5}$$

Hence, (10) becomes

$$\left.\frac{dw}{ds}\right|_{(1,2)} = (17,36) \cdot \left(\frac{3}{5},\frac{4}{5}\right)$$

$$= \frac{51}{5} + \frac{144}{5}$$

$$= \frac{196}{5} = 39.2$$

Notice how much more convenient this approach is than if we had to resort to the approach of part (a) every time we wanted to find a directional derivative.

d. Hopefully, we see at a glance from (10) that $\left.\frac{dw}{ds}\right|_{(1,2)}$ is maximum when $\vec{u}_s$ is in the direction of (17,36), and, in this event,

$$\left.\frac{dw}{ds}\right|_{(1,2)} = (17,36)\cdot\vec{u}_s = \|(17,36)\| \text{ [since } \vec{u}_s \text{ is parallel to (17,36), } \cos\angle\left((17,36), \vec{u}_s\right) = 1]$$

$$= \sqrt{17^2+36^2} = \sqrt{1585} \approx 39.8$$

3.3.4

a. Since $w = f(x,y) = e^x \cos y + x^2y$,

$$f_x(x,y) = e^x \cos y + 2xy$$

$$f_y(x,y) = -e^x \sin y + x^2$$

$$f_x(\ell n\ 2, \tfrac{\pi}{2}) = e^{\ell n 2} \cos \tfrac{\pi}{2} + (2\ \ell n\ 2)\ \tfrac{\pi}{2} = \pi\ \ell n\ 2$$

$$f_y(\ell n\ 2, \tfrac{\pi}{2}) = -e^{\ell n 2} \sin \tfrac{\pi}{2} + (\ell n 2)^2 = -2+(\ell n\ 2)^2$$

Therefore,

$$\vec{\nabla} f(\ell n 2, \tfrac{\pi}{2}) = [\pi\ \ell n\ 2, -2+(\ell n\ 2)^2]$$

b.

$$\left.\frac{dw}{ds}\right|_{(\ell n 2, \frac{\pi}{2})} = [\pi\ \ell n\ 2, -2+(\ell n 2)^2] \cdot (1,0)$$

$$= \pi \ell n\ 2$$

c.

$$\left.\frac{dw}{ds}\right|_{(\ell n 2, \frac{\pi}{2})} = [\pi\ \ell n\ 2, -2+(\ell n 2)^2] \cdot (0,1)$$

$$= -2+(\ell n 2)^2$$

d. The direction in which $\frac{dw}{ds}$ is maximum is in the direction of the gradient, which we computed in part (a) to be $[\pi\ \ell n\ 2, -2 + (\ell n\ 2)^2]$.

$$\max \left.\frac{dw}{ds}\right|_{(\ell n 2, \frac{\pi}{2})} = \|\vec{\nabla} f(\ell n 2, \tfrac{\pi}{2})\| = \sqrt{(\pi \ell n 2)^2 + [-2+(\ell n 2)^2]^2}$$

$$= \sqrt{\pi^2 \ell n^2 2 + 4 - 4 \ell n^2 2 + \ell n^4 2}$$

3.3.5(L)

The main aim of this exercise is to show that the definition of the gradient vector is independent of any particular coordinate system, but that the form of it does depend on the coordinate system.

In particular we wish to show that the convenient expression $f_x\vec{i} + f_y\vec{j}$ which expresses the gradient of f in terms of Cartesian coordinates is a rather special case. For example, if one were merely to memorize the expression for the gradient in Cartesian coordinates, one might expect that in, say, polar coordinates where the basic unit vectors are $\vec{u}_r$ and $\vec{u}_\theta$, the gradient would be given by $f_r\vec{u}_r + f_\theta\vec{u}_\theta$. The aim of this exercise is to show that this is not the case.

a. Given that r and θ are the independent variables of polar coordinates, we have from $w = r\sin\theta$ that $\frac{\partial w}{\partial r} = \sin\theta$ and $\frac{\partial w}{\partial \theta} = r\cos\theta$. Since $\vec{u}_r = \cos\theta\vec{i} + \sin\theta\vec{j}$ and $\vec{u}_\theta = -\sin\theta\vec{i}+\cos\theta\vec{j}$ (why?), it follows that

$$\frac{\partial w}{\partial r}\vec{u}_r + \frac{\partial w}{\partial \theta}\vec{u}_\theta$$

$$= \sin\theta(\cos\theta\vec{i}+\sin\theta\vec{j})+r\cos\theta(-\sin\theta\vec{i}+\cos\theta\vec{j})$$

$$= (\sin\theta\cos\theta - r\sin\theta\cos\theta)\vec{i}+(\sin^2\theta+r\cos^2\theta)\vec{j}.$$

Recalling that $\sin\theta = \frac{y}{r}$ and $\cos\theta = \frac{x}{r}$, we obtain

$$\frac{\partial w}{\partial r}\vec{u}_r + \frac{\partial w}{\partial \theta}\vec{u}_\theta = \left(\frac{xy}{r^2} - \frac{xy}{r}\right)\vec{i} + \left(\frac{y^2}{r^2} + \frac{x^2}{r}\right)\vec{j}$$

$$= \frac{xy}{r^2}(1-r)\vec{i} + \frac{1}{r^2}(y^2+rx^2)\vec{j}$$

$$= \frac{xy}{x^2+y^2}\left(1-\sqrt{x^2+y^2}\right)\vec{i} + \frac{1}{x^2+y^2}\left(y^2+x^2\sqrt{x^2+y^2}\right)\vec{j} \quad (1)$$

3.3.5(L) continued

b. If we convert to Cartesian coordinates, we have that $w = r \sin\theta$ implies

$$w = y \quad ,$$

whence

$$\frac{\partial w}{\partial x} = 0 \quad \text{and} \quad \frac{\partial w}{\partial y} = 1^{*}$$

Hence,

$$\frac{\partial w}{\partial x}\vec{i} + \frac{\partial w}{\partial y}\vec{j} = \vec{j} \qquad (2)$$

But we have already seen that $\vec{\nabla}w = \frac{\partial w}{\partial x}\vec{i} + \frac{\partial w}{\partial y}\vec{j}$.

Hence, from (2),

$$\vec{\nabla}w = \vec{j} \; .$$

While parts (a) and (b) should not have caused you too much difficulty to obtain, the implication we want to make is that we must not be spoiled by the convenience of Cartesian coordinates and conclude that if w is expressed in terms of u and v and if $\vec{e}_u$ and $\vec{e}_v$ represent the unit vectors in this new coordinate system, then $\frac{\partial w}{\partial u}\vec{e}_u + \frac{\partial w}{\partial v}\vec{e}_v$ represents the gradient of w with respect to u and v. To be sure we can compute the vector, but it need not be the gradient. This is precisely what we showed in this exercise. Namely, equation (1) made well-defined sense, but equation (2) showed that (1) was not the gradient vector.

With respect to this particular exercise, we may use the following "geo-analytic" approach. If we elect to write all quantities at a point in terms of $\vec{u}_r$ and $\vec{u}_\theta$ then

---

*See Note at the end of this exercise.

3.3.5(L) continued

$$\vec{\nabla}w = [\ ]\ \vec{u}_r + (\ )\ \vec{u}_\theta \tag{3}$$

We utilize the fact that $\vec{u}_r \cdot \vec{u}_\theta = 0$, to obtain from (3) that

$$\vec{\nabla}\ w \cdot \vec{u}_r = [\ ] \tag{4}$$

while

$$\vec{\nabla}w \cdot \vec{u}_\theta = (\ ) \tag{5}$$

But $\vec{\nabla}w\cdot\vec{u}_r$, by the definition of the gradient, is the directional derivative of w in the direction of $\vec{u}_r$, and since $\vec{u}_r$ is measured radially, it seems geometrically clear that this directional derivative is simply $\frac{\partial w}{\partial r}$.
That is,

$$[\ ] = \frac{\partial w}{\partial r}$$

In a similar manner, $\vec{\nabla}w\cdot\vec{u}_\theta$ is the derivative of w in the direction of $\vec{u}_\theta$. Now, $\vec{u}_\theta$ is not in the direction of θ, since (among other things) this doesn't mean anything. What is true is that the direction of $\vec{u}_\theta$ is that of a 90° positive rotation of $\vec{u}_r$, and for small $\Delta\theta$, the change at right angles to u_r is measured by $r\Delta\theta$. In other words, our geometric intuition might now lead us to suspect that the derivative of w in the direction of $\vec{u}_\theta$ is

$$\lim_{\Delta\theta\to 0} \frac{\Delta w}{r\Delta\theta} \text{ or, } \frac{1}{r}\ \frac{\partial w}{\partial \theta} .$$

In other words,

$$(\ ) = \frac{1}{r}\ \frac{\partial w}{\partial\theta}$$

Without worrying how to determine this result more rigorously (or for that matter, more generally in the case that the coordinate system is not as simple as Polar coordinates), it appears that a reasonable guess for w is

3.3.5(L) continued

$$\vec{\nabla}w = \frac{\partial w}{\partial r}\vec{u}_r + \frac{1}{r}\frac{\partial w}{\partial \theta}\vec{u}_\theta \qquad (6)$$

To conclude our discussion, let us compute the right side of (6) in our present exercise and see how it compares with the gradient found in part (b).

To this end:

We already know that

$$\frac{\partial w}{\partial r} = \sin\theta,\ \frac{\partial w}{\partial \theta} = r\cos\theta,\ \vec{u}_r = \cos\theta\vec{i}+\sin\theta\vec{j},\text{ and } \vec{u}_\theta = -\sin\theta\vec{i}+\cos\theta\vec{j}$$

Hence,

$$\frac{\partial w}{\partial r}\vec{u}_r + \frac{1}{r}\frac{\partial w}{\partial \theta}\vec{u}_\theta$$

$$= \sin\theta(\cos\theta\vec{i} + \sin\theta\vec{j}) + \cos\theta(-\sin\theta\vec{i} + \cos\theta\vec{j})$$

$$= (\sin^2\theta + \cos^2\theta)\ \vec{j} = \vec{j} \quad ,$$

and this agrees with $\vec{\nabla}w$ as determined in part (b) of this exercise.

Note:

Suppose we are given that

$$w = f(x,y) \qquad (1)$$

and that we also know that x and y are themselves functions of two other variables, say, u and v. For example,

$$x = g(u,v) \text{ and } y = h(u,v) \qquad (2)$$

Putting (2) into (1) yields

$$w = f(g(u,v),h(u,v)) \qquad (3)$$

and from (3) we see that w is now a function of u and v. (This discussion will ultimately generalize in a later unit to the chain rule for several variables, but this is not important at this moment.)

The point is that how u and v must be combined to yield a particular w need not be the same as how x and y must be combined to yield this same value of w. To make this point more concretely, suppose that

$$w = f(x,y) = x + y \tag{1'}$$

and that

$$x = u^2 \text{ while } y = -v^2 \tag{2'}$$

Putting (2') into (1') yields

$$w = f(u^2, -v^2) = u^2 - v^2 \tag{3'}$$

The point is that (1') and (3') define entirely different functions. Namely in (1') we add the first component to the second to obtain w, while in (3') we subtract the square of the second component from the square of the first to obtain w. In other words, we could have rewritten (3') as

$$w = f(u^2, -v^2) = u^2 - v^2 = F(u,v) \tag{3''}$$

where we use F to indicate that the rule for combining u and v to form w is different from the rule for combining x and y to yield w.

Returning to the general case, (3) may be written in the form

$$w = F(u,v) \tag{4}$$

If we now want to form the partial of w with respect to u, it is clear that we are referring to $F_u(u,v)$, while if we wish to form the partial of w with respect to x, it is clear that we are referring to $f_x(x,y)$.

With regard to the present exercise, when we write $\frac{\partial w}{\partial r}$ or $\frac{\partial w}{\partial \theta}$ we are referring to the formula $w = r \sin \theta$ while if we write $\frac{\partial w}{\partial x}$ or $\frac{\partial w}{\partial y}$ we are referring to the formula $w = y$.

3.3.6(L)

This exercise is actually a buffer (in terms of its geometric simplicity) to help us clarify a point that is often difficult for the beginner to grasp. Given $w = f(x,y)$, we have seen graphically that the gradient is a vector in the xy-plane which points in the direction of the maximum $\frac{dw}{ds}$ and whose magnitude is this maximum value. Now "all of a sudden" (in particular, see Exercises 3.3.7 and 3.3.8) we are talking about the gradient not being in the xy-plane but rather being normal to our surface. There is a subtle (but extremely inportant) difference between the role of $\vec{\nabla}f$ when we think of f in the form, $w = f(x,y,z)$; or in the form $f(x,y,z) = c$.

To see this difference as clearly as possible, it was our feeling that we should introduce the idea with an example involving an even lower dimension than that discussed in the text.

We have that $w = f(x,y) = y - x^2$.

From this it is easy to see that $\vec{\nabla}f$ is given by $-2x\vec{1} + \vec{j}$. This tells us that if we move from the point (x_0,y_0) the maximum directional derivative, $\frac{dw}{ds}$, is

$$\sqrt{4x_0^2 + 1}\quad,$$

and this occurs in the direction $-2x_0\vec{1} + \vec{j}$.

So far, so good! Now, we consider a completely different question. We look at a special set of points (x,y) in the xy-plane for which $f(x,y) = 4$; i.e., $y - x^2 = 4$. In this event, x and y are no longer independent variables since the restriction that $f(x,y)=4$, or in this case, $y-x^2=4$, fixes the value of y, for example, once the value of x is known.

3.3.6(L) continued

While it is clear that $y-x^2 = 4$ is a curve in the xy-plane, relative to $w = y-x^2$, this curve consists precisely of those points for which w=4. In other words, we say that $y-x^2 = 4$ is an <u>equipotential curve</u> of the <u>surface</u>, $w = y-x^2$. If we let C denote the curve $y-x^2 = 4$, it follows that $\frac{dw}{ds}$ must be identically zero, where s denotes the direction of C (i.e., s denotes the direction of the line tangent to C at any point), since w is constant along C.

But we know that

$$\frac{dw}{ds} = \vec{\nabla}f \cdot \vec{u}_s$$

where $\vec{u}_s$ is the unit tangent vector C at any given point. Since $\frac{dw}{ds} = 0$, it must be that either $\vec{\nabla}f$ is $\vec{0}$ or that $\vec{\nabla}f$ is perpendicular to $\vec{u}_s$, since these are the only ways in which $\vec{\nabla}f \cdot \vec{u}_s$ can equal zero (since $\vec{u}_s$ is a unit vector, it can't equal $\vec{0}$). Now we have already seen that $\vec{\nabla}f = -2x\,\vec{1}+\vec{j}$; and since the coefficient of j is always 1, we see that $\vec{\nabla}f$ is never equal to $\vec{0}$. Hence, it follows that $\vec{\nabla}f$ is perpendicular to C.

As a quick check, we may observe that the slope of $y = x^2-4$ at (x_o,y_o) is given by $2x_o$, while the slope of $-2x\vec{1}+\vec{j}$ at this same point is $1/-2x_o$. Since $2x_o$ and $1/-2x_o$ are negative reciporcals, we see that the gradient of f at (x_o,y_o) gives the slope of the line normal to $f(x,y) = 4$ at this point.

In a similar way, when we work with $w = f(x,y,z)$, the gradient of f is a vector in 3-dimensional space whose direction and magnitude tell us the direction in which $\frac{dw}{ds}$ is a maximum and what the value of this maximum is. When, instead, we work with the <u>surface</u> (2 degrees of freedom) $f(x,y,z) = $ constant, then the gradient of f is a vector normal to this surface. The key is to distinguish between $w = f(x,y,z)$ and $f(x,y,z) = c$. In this particular exercise we compared $w = f(x,y)$ with $f(x,y) = $ constant.

3.3.6(L) continued

As a final note on this exercise, we should like to discuss the notion of orthogonal trajectories. Suppose we are given a 1-parameter family of curves (i.e., the family contains one arbitrary constant). The most general representation of such a family is

$$f(x,y,c) = 0 .$$

(That is, there is some functional relationship between x, y, and an arbitrary constant, c).

We call the family of curves $g(x,y,c) = 0$ a family of orthogonal trajectories to the family $f(x,y,c) = 0$ if at every point of intersection between a member of $g(x,y,c) = 0$ and a member of $f(x,y,c) = 0$, the angle of intersection is 90° (hence the name orthogonal). Recall that angles between curves are defined to be the angles between their tangent lines at the point of intersection.

Computationally, the way we find a family of orthogonal trajectories (if indeed such a family exists), is that we compute $\frac{dy}{dx}$ from $f(x,y,c) = 0$ in a way that eliminates c. That is, we express $\frac{dy}{dx}$ in the form $\frac{dy}{dx} = h(x,y)$. We then solve the differential equation

$$\frac{dy}{dx} = \frac{-1}{h(x,y)} ,$$

the idea being that the resulting curves must be orthogonal to those of the first family at the points of intersection by virtue of the fact that their slopes are negative reciprocals.

In our particular example, we have that the 1-parameter family of curves $y-x^2 = c$ (which are the equipotential curves for $w = y-x^2$) satisfies the differential equation

$$\frac{dy}{dx} = 2x .$$

3.3.6(L) continued

Hence the family of orthogonal trajectories is given by the differential equation

$$\frac{dy}{dx} = \frac{-1}{2x}$$

Solving this equation yeilds

$$dy = \frac{-dx}{2x}$$

or

$$y = -\frac{1}{2}\ \ell n|x| + c$$

Hopefully, it is clear that the notion of orthogonal trajectories is not restricted to gradients and equipotential curves, but there is an interesting application in this respect. Suppose we are given the surface $w = f(x,y)$ and we wish to start at the point (x_0,y_0) and move in such a way in the xy-plane that the height of the surface is always changing as rapidly as possible.

What we do is sketch the equipotential curves of w. We then start at (x_0,y_0) on the equipotential curve $w_0 = f(x,y)$, where w_0 is the value of $f(x_0,y_0)$, and we move along the orthogonal trajectory to this equipotential curve. In other words (and this may seem clear from an intuitive point of view), to make $\frac{dw}{ds}$ maximum we must "cut across" the family of equipotential curves as swiftly as possible, and this is done by moving orthogonally to the family.

At any rate, it is our hope that this discussion has cleared up the distinction between the role played by $\vec{\nabla}f$ when we consider $w = f(x,y)$ and the role of $\vec{\nabla}f$ when we consider $f(x,y) = c$. The next two exercises extend this notation to $w = f(x,y,z)$ and $f(x,y,z) = 0$.

3.3.7(L)

In our previous encounters with finding normals to surfaces, our surface was in the form $z = f(x,y)$. In the present exercise, it is at best difficult to solve for any variable explicitly in terms of the other two.

Now, given

$$x^5+y^4z^3+xyz^5 = 3 \tag{1}$$

we may think of a new function

$$g(x,y,z) = x^5+y^4z^3+xyz^5 \tag{2}$$

In (2), x, y, z are independent variables. Again, since g is a polynomial in x, y, and z, we may compute the gradient of g to obtain

$$\vec{\nabla}g = (5x^4+yz^5,\ 4y^3z^3+xz^5,\ 3y^4z^2+5xyz^4)$$

therefore,

$$\vec{\nabla}g(1,1,1) = (6,5,8) \tag{3}$$

Now, equation (1) may be viewed as a special set of points, S, for which $g(x,y,z) = 3$ for all (x,y,z) S. In other words, restricted to S, g is a constant. Hence, on S, $\Delta g = 0$.

But $\frac{\Delta g}{\Delta s} = \vec{\nabla}g \cdot \vec{u}_s$ along any curve s in the surface S. That is $\vec{\nabla}g \cdot \vec{u}_s$ must be zero; yet, from (3) $\vec{\nabla}g \neq \vec{0}$, nor can $\vec{u}_s = \vec{0}$ since $\vec{u}_s$ is a unit vector.

Hence, $\vec{\nabla}g \cdot \vec{u}_s = 0 \rightarrow \vec{\nabla}g \perp \vec{u}_s$.

Since $\vec{\nabla}g$ is perpendicular to each curve on S at (1,1,1), $\vec{\nabla}g$ is normal to S.

In other words

$6\vec{1} + 5\vec{j} + 8\vec{k}$

is a vector normal to the surface $x^5+y^4z^3+xyz^5 = 3$ at the point (1,1,1)

A Note on Normal Lines to Surfaces

Suppose we are given the surface whose equation is, say,

$$f(x,y,z) = c \quad . \tag{1}$$

There is a danger that you may confuse two quite different points here. For the sake of argument, let us suppose that equation (1) can be solved for z, explicitly in terms of x and y. That is, let us suppose that the surface S whose equation is given by equation (1), is also given by

$$z = g(x,y) \tag{2}$$

It is important to note that it is the gradient of f, not g, that is normal to the surface S. In fact, the gradient of g is a vector in the xy-plane, and it is hardly likely that the normal to the surface S would always be in the xy-plane (or parallel to it).

To see this from a different perspective, suppose we start with equation (2). We could then define a function f(x,y,z) by rewriting (2) as

$$z-g(x,y) = 0$$

and let

$$f(x,y,z) = z-g(x,y) \quad . \tag{3}$$

In this case, $f(x,y,z) = 0$. That is, the gradient of f is normal to the surface defined by (2) since this surface is an equipotential surface of f(x,y,z).

As a check, notice that the gradient of f in this case is the vector

$$-g_x\vec{i}-g_y\vec{j}+\vec{k} \tag{4}$$

and this checks with our earlier result that the tangent plane to the surface $z = g(x,y)$ has the equation:

$$z-z_0 = g_x(x_o,y_o)(x-x_0)+g_y(x_o,y_o)(y-y_o) \quad ,$$

from which we can see that the normal is indeed as given by (4).

3.3.8

We have

$$f(x,y,z) = x^4+y^6z+xyz^5$$

Hence,

$$\vec{\nabla}f = (4x^3+yz^5,\ 6y^5z+xz^5,\ y^6+5xyz^4)$$

therefore

$$\vec{\nabla}f(1,1,1) = (5,7,6) \qquad (1)$$

Since $x^4+y^6z+xyz^5 = 3$ defines an equipotential surface of f (i.e., $\Delta f = 0$ on S), we have from (1) that $5\vec{i}+7\vec{j}+6\vec{k}$ is normal to S at (1,1,1).

Therefore the equation of the tangent plane is

$$5(x-1)+7(y-1)+6(z-1) = 0$$

or

$$5x+7y+6z = 18 \quad .$$

(Notice that the use of the normal and the equation of a plane are the same as always; all that's new in the last two problems is the technique for finding the normal to a surface.)

3.3.9

a. We simply invoke the technique of writing ab+cd as $(a,c)\cdot(b,d)$. This yields

$$\Delta f = [f_{x_1}(a_1,a_2),f_{x_2}(a_1,a_2)]\cdot(\Delta x_1,\Delta x_2)+(k_1,k_2)\cdot(\Delta x_1,\Delta x_2) \qquad (1)$$

If we now let $\underline{k} = (k_1,k_2)$, $\Delta\underline{x} = (\Delta x_1,\Delta x_2)$ and observe that $[f_{x_1}(a_1,a_2),f_{x_2}(a_1,a_2)]$ is $\nabla f(\underline{a})$, we obtain

$$\Delta f = \nabla f(\underline{a})\cdot\Delta\underline{x} + \underline{k}\cdot\Delta\underline{x} \qquad (2)$$

where $\lim_{\Delta\underline{x}\to\underline{0}} \underline{k} = \underline{0}$.

3.3.9 continued

The validity of (2) required that f_{x_1} and f_{x_2} exist and be continuous at $\underline{x} = \underline{a}$. Once these requirements are met we observe that equation (2) bears a strong resemblance to the 1-dimensional result that if f is differentiable at x = a, then

$$\Delta f = f'(a)\Delta x + k\Delta x, \text{ where } \lim_{\Delta x \to 0} k = 0 \qquad (3)$$

The interesting point is that (3) is a special case of (2). That is, if we identify $\nabla f(\underline{a})$ with the n-dimensional analog of f'(a) and realize that in 1-dimensional space the dot product and the "ordinary product" are the same, we see that when translated into 1-space equation (2) becomes equation (3).

More generally the most important aspect of (2) is that it applies in every dimension. Granted that our text proved it in the 2-dimensional case and indicated that the proof applied almost verbatim in higher dimensions, the fact is that the text was conducive to our concluding that there were geometrical reasons for (2) being true. Our aim now is to show that the technique of the text does apply virtually word-for-word to higher dimensions in which there is no intuitive appeal to geometry. We have chosen the case n = 4. The point is that our proof reviews the one in the text as we show that it applies to all dimensions. Admittedly, the proof is abstract but we feel that it is important for you to suffer through it in the hope that you will once and for all get the true feeling of what the result means.

b. We have that $f: E^4 \to E$ is continuous and that $\underline{a} = (a_1, a_2, a_3, a_4) \in E^4$. We are also told that $f_{x_1}, f_{x_2}, f_{x_3}$, and f_{x_4} all exist and are continuous at $\underline{a}$. We want to show that

3.3.9 continued

$$\Delta f = [f(a_1+\Delta x_1, a_2+\Delta x_2, a_3+\Delta x_3, a_4+\Delta x_4) - f(a_1,a_2,a_3,a_4)]^*$$

$$= \vec{\nabla} f(a_1,a_2,a_3,a_4)\cdot\underline{\Delta x}+\underline{k}\cdot\underline{\Delta x}$$

where $\lim_{\underline{\Delta x}\to\underline{0}} \underline{k} = \underline{0}$

The "trick" in problems of this type is to add and subtract terms in such a way that in each difference all but one of the variables are being held constant. All that is required to make sure that we can do this is that the variables be <u>independent</u>.

The only other "trick" is that there are many different ways that we can add and subtract the terms. The key point is that by continuity, we need not worry about more than one particular path, since continuity guarantees us that whatever answer we get along one path we get along any other path as well.

Thus, for example, we may write

$$\left.\begin{aligned}\Delta f &= f(a_1+\Delta x_1, a_2+\Delta x_2, a_3+\Delta x_3, a_4+\Delta x_4) - f(a_1,a_2,a_3,a_4)\\ &= [f(a_1+\Delta x_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)]\\ &+ [f(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2,a_3+\Delta x_3,a_4+\Delta x_4)]\\ &+ [f(a_1,a_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2,a_3,a_4+\Delta x_4)]\\ &+ [f(a_1,a_2,a_3,a_4+\Delta x_4)-f(a_1,a_2,a_3,a_4)]\end{aligned}\right\} (4)$$

---

*In n-tuple notation,

$\Delta f = f(\underline{a}+\underline{\Delta x})-f(\underline{a})$

where $\underline{\Delta x} = (\Delta x_1,\Delta x_2,\Delta x_3,\Delta x_4)$

3.3.9 continued

Equation (4) may then be rewritten so that the various partial derivatives are emphasized (and this is why we assume that f_{x_1}, f_{x_2}, f_{x_3}, and f_{x_4} exist in a neighborhood of $\underline{a}$, for we can't really take advantage of the partial derivatives if they don't exist!)

We then obtain from (4) that

$$\begin{aligned}\Delta f = &\left[\frac{f(a_1+\Delta x_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)}{\Delta x_1}\right]\Delta x_1\\ &+\left[\frac{f(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2,a_3+\Delta x_3,a_4+\Delta x_4)}{\Delta x_2}\right]\Delta x_2\\ &+\left[\frac{f(a_1,a_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2,a_3,a_4+\Delta x_4)}{\Delta x_3}\right]\Delta x_3\\ &+\left[\frac{f(a_1,a_2,a_3,a_4+\Delta x_4)-f(a_1,a_2,a_3,a_4)}{\Delta x_4}\right]\Delta x_4\end{aligned} \qquad (5)$$

The first bracketed expression in (5) is related to f_{x_1} in the sense that

$$f_{x_1}(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)$$

$$= \lim_{\Delta x_1\to 0}\left[\frac{f(a_1+\Delta x_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)-f(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)}{\Delta x_1}\right]$$

Thus for a fixed $\Delta x_1 \neq 0$, the definition of $\lim_{\Delta x_1\to 0}$ implies that

3.3.9 continued

$$\left[\frac{f(a_1+\Delta x_1, a_2+\Delta x_2, a_3+\Delta x_3, a_4+\Delta x_4) - f(a_1, a_2+\Delta x_2, a_3+\Delta x_3, a_4+\Delta x_4)}{\Delta x_1}\right]$$

$$= f_{x_1}(a_1, a_2+\Delta x_2, a_3+\Delta x_3, a_4+\Delta x_4) + \varepsilon_1$$

where $\lim\limits_{\Delta x_1 \to 0} \varepsilon_1 = 0$.

Using similar reasoning we may also obtain

$$\left[\frac{f(a_1, a_2+\Delta x_2, a_3+\Delta x_3, a_4+\Delta x_4) - f(a_1, a_2, a_3+\Delta x_3, a_4+\Delta x_4)}{\Delta x_2}\right]$$

$$= f_{x_2}(a_1, a_2, a_3+\Delta x_3, a_4+\Delta x_4) + \varepsilon_2$$

$$\left[\frac{f(a_1, a_2, a_3+\Delta x_3, a_4+\Delta x_4) - f(a_1, a_2, a_3, a_4+\Delta x_4)}{\Delta x_3}\right]$$

$$= f_{x_3}(a_1, a_2, a_3, a_4+\Delta x_4) + \varepsilon_3$$

$$\left[\frac{f(a_1, a_2, a_3, a_4+\Delta x_4) - f(a_1, a_2, a_3, a_4)}{\Delta x_4}\right] = f_{x_4}(a_1, a_2, a_3, a_4) + \varepsilon_4$$

where ε_2, ε_3, $\varepsilon_4 \to 0$ as Δx_2, Δx_3, $\Delta x_4 \to 0$.

Thus (5) becomes

3.3.9 continued

$$\left.\begin{aligned}\Delta f = &f_{x_1}(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4)\Delta x_1+\varepsilon_1\Delta x_1\\ &+ f_{x_2}(a_1,a_2,a_3+\Delta x_3,a_4+\Delta x_4)\Delta x_2+\varepsilon_2\Delta x_2\\ &+ f_{x_3}(a_1,a_2,a_3,a_4+\Delta x_4)\Delta x_3+\varepsilon_3\Delta x_3\\ &+ f_{x_4}(a_1,a_2,a_3,a_4)\Delta x_4+\varepsilon_4\Delta x_4\end{aligned}\right\} \quad (6)$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \to 0$ as $\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_4 \to 0$.

All that keeps (6) from being the correct answer are the points at which $f_{x_1}, f_{x_2}, f_{x_3}$ are computed. To obtain the desired result, we need only assume that $f_{x_1}, f_{x_2}, f_{x_3}$, and f_{x_4} are continuous at $\underline{a} = (a_1,a_2,a_3,a_4)$. [Actually as we are doing the problem it appears that we have no worries about f_{x_4} since it is being evaluated at the desired point $\underline{a}$. However, as we mentioned we could have chosen other paths in which case f_{x_4} might have been evaluated at a different point. The assumption, therefore, that f_{x_4} is continuous at $\underline{a}$ protects us in the event we wish to use a different path.]

Namely, for example, the fact that f_{x_1} is continuous at $\underline{a}$ means

$$f_{x_1}(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4) \to f_{x_1}(a_1,a_2,a_3,a_4)$$

as $(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4) \to (a_1,a_2,a_3,a_4)$.

3.3.9 continued

That is,

$$f_{x_1}(a_1,a_2+\Delta x_2,a_3+\Delta x_3,a_4+\Delta x_4) = f_{x_1}(a_1,a_2,a_3,a_4) + c_1$$

where $c_1 \to 0$ as $\Delta x_2,\ \Delta x_3,\ \Delta x_4 \to 0$.

Similarly,

$$f_{x_2}(a_1,a_2,a_3+\Delta x_3,a_4+\Delta x_4) = f_{x_2}(a_1,a_2,a_3,a_4) + c_2$$

where $c_2 \to 0$ as $\Delta x_3,\ \Delta x_4 \to 0$

$$f_{x_3}(a_1,a_2,a_3,a_4+\Delta x_4) = f_{x_3}(a_1,a_2,a_3,a_4) + c_3$$

where $c_3 \to 0$ as $\Delta x_4 \to 0$

whereupon (6) becomes

$$\left.\begin{aligned}\Delta f = {} & f_{x_1}(a_1,a_2,a_3,a_4)\Delta x_1 + (\varepsilon_1+c_1)\Delta x_1 \\ & + f_{x_2}(a_1,a_2,a_3,a_4)\Delta x_2 + (\varepsilon_2+c_2)\Delta x_2 \\ & + f_{x_3}(a_1,a_2,a_3,a_4)\Delta x_3 + (\varepsilon_3+c_3)\Delta x_3 \\ & + f_{x_4}(a_1,a_2,a_3,a_4)\Delta x_3 + \varepsilon_4\Delta x_4\end{aligned}\right\} \qquad (7)$$

Letting $k_1 = \varepsilon_1+c_1$, $k_2 = \varepsilon_2+c_2$, $k_3 = \varepsilon_3+c_3$, $k_4 = \varepsilon_4$, and abbreviating (a_1,a_2,a_3,a_4) by $\underline{a}$, (7) becomes

$$\left.\begin{aligned}\Delta f = {} & f_{x_1}(\underline{a})\Delta x_1+f_{x_2}(\underline{a})\Delta x_2+f_{x_3}(\underline{a})\Delta x_3+f_{x_4}(\underline{a})\Delta x_4 \\ & + [k_1\Delta x_1+k_2\Delta x_2+k_3\Delta x_3+k_4\Delta x_4] \\ & \text{where } k_1,k_2,k_3,k_4 \to 0 \text{ as } \Delta x_1,\ \Delta x_2,\ \Delta x_3,\ \Delta x_4 \to 0.\end{aligned}\right\} \qquad (8)$$

3.3.9 continued

We are now home free if we introduce the notation

$$\Delta\underline{x} = (\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_4), \quad \vec{\nabla} f(\underline{a}) = [f_{x_1}(\underline{a}), f_{x_2}(\underline{a}), f_{x_3}(\underline{a}), f_{x_4}(\underline{a})]$$

and

$$\underline{k} = (k_1, k_2, k_3, k_4),$$

for, then, (8) becomes

$$\Delta f = \vec{\nabla} f(\underline{a}) \cdot \Delta\underline{x} + \underline{k} \cdot \Delta\underline{x} \ , \quad \text{where} \lim_{\Delta\underline{x} \to \underline{0}} \underline{k} = \underline{0}$$

As a final note on this exercise, notice that we can always define the gradient of f at $(\underline{a}) = (a_1, a_2, a_3, a_4)$ to be the 4-tuple $[f_{x_1}(\underline{a}), f_{x_2}(\underline{a}), f_{x_3}(\underline{a}), f_{x_4}(\underline{a})]$, but if we want the difference between Δf and $\nabla f(\underline{a}) \cdot \Delta\underline{x}$ to go to zero faster than $\|\Delta x\|$, then we must insist that $f_{x_1}, f_{x_2}, f_{x_3}$, and f_{x_4} also be continuous at $\underline{a}$. Since we shall often require that the above difference does approach zero "sufficiently fast," we shall agree that even though the 4-tuple $[f_{x_1}(\underline{a}), f_{x_2}(\underline{a}), f_{x_3}(\underline{a}), f_{x_4}(\underline{a})]$ exists once the partial derivatives exist at $\underline{a}$, we will not call this 4-tuple the gradient of f at $\underline{a}$ unless the partials also happen to be continuous at $\underline{a}$.