5

DERIVATIVES IN N-DIMENSIONAL VECTOR SPACES

A

Introduction

The approach of our textbook (as developed in the previous two units) in emphasizing the geometrical aspects of partial derivatives, directional derivatives, and the gradient is, in many ways, an excellent approach. We get a picture of what appears to be happening, and, especially for the beginning student, this is much more meaningful than a more precise and general discussion that would apply to all vector spaces, regardless of dimension. In fact, if we want to use the latter approach, it is wise to use the former first. This is in accord with many such decisions which have already been made in our previous treatment of mathematics. As an example, within the framework of our present course, notice how carefully we exploited the geometric aspects of vectors as arrows before we even mentioned n-dimensional vector spaces. And, in this respect, recall how sophisticated many of our "arrow" results were in their own right. That is, granted that arrows were simpler than n-tuples, when all we had were arrows, there were still some rather difficult ideas for us to assimilate.

In any event, our aim in this chapter is to continue within the spirit of the game of mathematics which, thus far, has served us so well. Namely, we have introduced n-dimensional vector spaces so that we could utilize the similarity in form between the previously studied functions of the form $f(x)$ and our new functions of the form $f(\underline{x})$. For example, when it came time to define continuity and limits for functions of several real variables, we rewrote $f(x_1,\ldots,x_n)$ as $f(\underline{x})$ and then mimicked the $f(x)$ situation by copying definitions verbatim, subject only to making the appropriate "vectorizations."

Rather than review this procedure, let us instead move to the next plateau. Using the approach developed in Chapter 4 of these supplementary notes, once we had defined what limits and continuity meant for any function, $f:E^n\rightarrow E$, our next logical step would have been to mimic our definition of $f'(a)$, and thus "induce" a meaning of $f'(\underline{a})$ where $\underline{a}\in E^n$.

Our aim in this chapter is to see where such an attempt would have led us and, once this is done, to see how our "new" concept of differentiation is related to the more traditional concept of differentiation as it was presented in the previous two units.

B

A New Type of Quotient

From the definition

$$f'(a) = \lim_{\Delta x \to 0} \left[\frac{f(a+\Delta x)-f(a)}{\Delta x} \right] \quad (1)$$

it seems natural that if now $f:E^n \to E$ and $\underline{a} \in E^n$, we should define $f'(\underline{a})$ by

$$f'(\underline{a}) = \lim_{\Delta \underline{x} \to \underline{0}} \left[\frac{f(\underline{a}+\Delta \underline{x})-f(\underline{a})}{\Delta \underline{x}} \right] , \quad (2)$$

where (2) is obtained from (1) by the appropriate "vectorizations."

In arriving at (2), the most subtle point is to observe that we must replace Δx by $\Delta \underline{x}$, since, among other things, $\underline{a}+\Delta x$ has no meaning (i.e., we have no rule for adding a vector [$\underline{a}$] to a real number [Δx]). Once Δx is replace by $\Delta \underline{x}$, it is clear that 0 must be replaced by $\underline{0}$.

If we next look at the bracketed expression in (2), we get a rather elegant insight as to how new mathematical concepts are often born. For, while at first glance, the derivation of (2) from (1) may seem harmless, a second glance shows us that we have "invented" an operation which we have never encountered before in our study of vectors (either as arrows or as n-tuples). More specifically, in the bracketed expression, our numerator is a real number (i.e., while $\underline{a}$ is a vector [n-tuple], recall that our notation $f(\underline{a})$ indicates that our "output" is a number), and our denominator is a vector!

Thus, whether we like it or not, if we decide to accept (2) as a starting point, we are obliged to investigate the meaning of $\frac{c}{\underline{v}}$

where c denotes an arbitrary but fixed real number and $\underline{v}$ an arbitrary but fixed vector. Notice, of course, that we could elect to abandon (2), but, in the usual spirit of things, it hardly seems wise to abandon an approach that has been fruitful simply because one potential problem arrises. Rather, it seems wiser for us to try, at least, to find a practical definition of what it means to divide a number by a vector.

Once we have agreed on this course of action, we again fall back on our previous experience and recall how we defined division in the case of two numbers. We defined $\frac{a}{b}$ to be that number which when multiplied by b yielded a. In terms of a more computational form, $\frac{a}{b}$ was defined by

$$b \text{ "times" } \frac{a}{b} = a \tag{3}$$

With this in mind, it seems that a very natural way to mimic (3) is to define $\frac{c}{\underline{v}}$ by

$$\underline{v} \text{ "times" } \frac{c}{\underline{v}} = c \tag{4}$$

This immediately suggests that unless we want to invent new forms of multiplication, (and nothing precludes this possibility, except that we already have enough problems) the "times" in (4) must denote the dot product, for according to (4) we must multiply a vector ($\underline{v}$) by "something" to obtain a number, and of the types of multiplication we have discussed, only the dot product allows us to multiply a vector by "something" (in fact, by another vector) to obtain a number. Thus, the "times" in (4) must denote a dot product. And, it therefore appears that $\frac{c}{\underline{v}}$ must, itself, be a vector, since, as we have just said, the dot product combines two vectors to produce a number.

There is, however, a little complication that presents itself here. The problem actually existed when we talked about the quotient of two numbers, but, except in the rather special case in which the denominator was zero, the problem never occurred. The point we are driving at is that when we say "$\frac{a}{b}$ is the number which when multiplied by b yields a," what gives us the right to say "the?" Why can't there be more than one such number, or for that matter, why not no such number?

The language of sets provides us with a nice explanation. Suppose we define $\frac{a}{b}$ to be the <u>set</u> of all numbers, c, such that bxc = a. That is,

$$\frac{a}{b} = \{c : b \times c = a\} \qquad (5)$$

Then, <u>unless b = 0*</u>, the <u>set</u> $\frac{a}{b}$ consists of the <u>single number</u> a÷b (e.g., $\frac{6}{3} = \{c:3c=6\} = \{6 \div 3\} = \{2\}$), and hence there is no harm in confusing the <u>set</u> $\frac{a}{b}$ with the <u>number</u> $\frac{a}{b}$, where in the latter context, $\frac{a}{b}$ is used as the fraction which denotes a÷b.

Mimicking (5) we may also define $\frac{c}{\underline{v}}$ to be a set; namely

$$\frac{c}{\underline{v}} = \{\underline{x} : \underline{x} \cdot \underline{v} = c\} \qquad (6)$$

The problem is that, unless $\underline{v} = \underline{0}$ and $c \neq 0$ (in which case $\frac{c}{\underline{v}} = \emptyset$), the set $\frac{c}{\underline{v}}$ has <u>infinitely many</u> elements!

Before we document this last remark, let us make sure that it is clear to you that the set described in (6) is a well-defined set <u>regardless of the dimension of the vector space</u>. In other words, recall that in the last chapter of these supplementary notes, we generalized the definition of a dot product so that it existed in any dimensional space. By way of a brief review, if $\underline{x} = (x_1,\ldots,x_n)$ and $\underline{y} = (y_1,\ldots,y_n)$ then $\underline{x} \cdot \underline{y} = x_1y_1 + \ldots + x_ny_n$.

Thus, by way of an example, suppose c = 5 and $\underline{v} = (1,2,3,4) \epsilon E^4$. Then by (6),

$$\frac{5}{(1,2,3,4)} = \{\underline{x} \epsilon E^4 : (1,2,3,4) \cdot \underline{x} = 5\} \qquad (7)$$

*If b=0, we have seen that if $a \neq 0$ then $\frac{a}{b} = \emptyset$ since no number times 0 can yield a non-zero number, and if b=0 and a=0, we have seen that $\frac{a}{b}$ (i.e., $\frac{0}{0}$) is the set of <u>all</u> numbers, since any number times zero is zero.

Notice that while it may be difficult to think of $\frac{5}{(1,2,3,4)}$ pictorially, (7) shows us that it is a "very real" set, at least in the sense that it is the solution set of the "very real" linear equation in four unknowns

$$x_1 + 2x_2 + 3x_3 + 4x_4 = 5 \qquad (8)$$

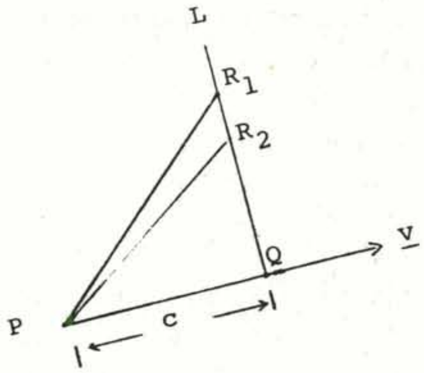
In fact, it might seem more natural to you if we restated this last remark in sort of a "reverse" way. Suppose in a traditional math course we were asked to find all solutions of the equation given by (8). Among other things, we may pick values for x_2, x_3, and x_4 completely at random, and once these random choices are made, x_1 can then be uniquely determined from (8) simply by letting $x_1 = 5 - 2x_2 - 3x_3 - 4x_4$. In modern language, this means that the solution set of (8) has infinitely many elements, and, by (7), another name for this infinite solution set is $\frac{5}{(1,2,3,4)}$.

In still other words, while we may not be used to thinking of the solutions of an equation like (8) as being points in a 4-dimensional vector space, the fact is that, conceptually, the idea is sound. We admit, however, that in the spirit of the text, there is probably more satisfaction if we think of the special cases in which we may view our vectors (n-tuples) as arrows and see what the geometric implications are. For the sake of a bit of simplicity (and we shall show in a moment that this restriction in no way loses any generality), let us assume that in (6), $\underline{v}$ denotes a unit vector (because if $\underline{v}$ is a unit vector, $\underline{x}\cdot\underline{v}$ is simply the projection of $\underline{x}$ in the direction of $\underline{v}$ while if $\underline{v}$ is not a unit vector, we must merely come to grips with the more computational fact that $\underline{x}\cdot\underline{v}$ is a scalar multiple of the projection of $\underline{x}$ in the direction of $\underline{v}$).

So, under the assumption that $\underline{v}$ is a unit vector and that our vectors are now arrows (in the diagrams which follow, we view our vectors as planar arrows), we may view

$$\frac{c}{\underline{v}} = \{\underline{x}:\underline{x}\cdot\underline{v} = c\}$$

as the set of all vectors whose projection in the direction of $\underline{v}$ is c. Clearly, there are infinitely many such vectors and this is amplified in Figure 1.



2. Thus $\vec{PR}_1$, $\vec{PR}_2$ etc. all have the property that $\underline{v}\cdot\vec{PR}_n = c$, when R_n is any point on L.

3. Therefore, according to (6), any vector of the form $\vec{PR}_n$ is an acceptable value of $\frac{c}{\underline{v}}$, i.e., $\underline{v}\cdot\vec{PR}_n = c$.

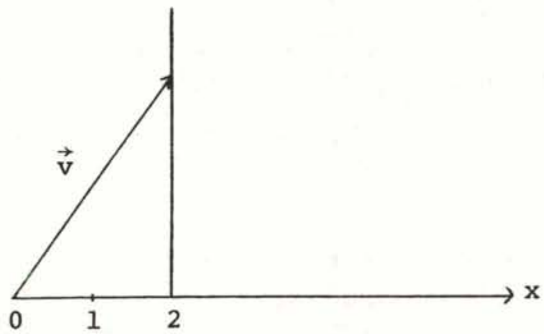
1. Let $|\vec{PQ}| = c$. Then if R is any point on the line L through Q perpendicular to $\underline{v}$, $\vec{PR}\cdot\underline{v} = c$, since $\underline{v}$ is a unit vector.

(Figure 1)

To see Figure 1 in terms of a more specific example, consider the set defined by

$$\frac{2}{\vec{i}} = \{\vec{v}: \vec{v}\cdot\vec{i} = 2\}$$

But $\vec{v}\cdot\vec{i}$ is the projection of $\vec{v}$ in the direction of the x-axis. Thus,



For any such $\vec{v}$, $\vec{v}\cdot\vec{i} = 2$; hence, any such $\vec{v}$ qualifies as a member of $\frac{2}{\vec{i}}$.

(Figure 2)

Our results are not seriously affected if $\underline{v}$ is not a unit vector. The key lies in the fact that if $\underline{v} \neq \underline{0}$, $\frac{\underline{v}}{\|\underline{v}\|}$ is a unit vector* in the direction of $\underline{v}$. In this event we have:

$$\frac{c}{\underline{v}} = \{\underline{x}: \underline{x}\cdot\underline{v} = c\} = \left\{\underline{x}: \underline{x}\cdot\frac{\underline{v}}{\|\underline{v}\|} = \frac{c}{\|\underline{v}\|}\right\} \tag{9}$$

Thus, from (9) we see that in general, for $\underline{v} \neq \underline{0}$, $\frac{c}{\underline{v}}$ is the set of all vectors whose projection in the direction of $\underline{v}$ is $\frac{c}{\|\underline{v}\|}$, and in the special case that $\underline{v}$ is a unit vector, $\|\underline{v}\|=1$, and $\frac{c}{\|\underline{v}\|}$ is simply c.

In the event $\underline{v} = \underline{0}$, $\|\underline{v}\|=0$ and accordingly $\frac{c}{\|\underline{v}\|} = \frac{c}{0}$, which is the "forbidden" quotient of two numbers. As a small aside, observe that if $\underline{v} = \underline{0}$, $\frac{c}{\underline{v}}$ behaves as its numeral counterpart. Namely,

$\frac{c}{\underline{0}} = \{\underline{x}: \underline{x}\cdot\underline{0} = c\}$. But $\underline{x}\cdot\underline{0} = 0$ [i.e., $(x_1,\ldots x_n)\cdot(0,\ldots,0) = x_1 0+\ldots+x_n 0 = 0+\ldots+0 = 0$]; hence, if $c \neq 0$, $\frac{c}{\underline{0}} = \emptyset$, while if $c = 0$, $\frac{c}{\underline{0}}$ is the entire vector space.

Now, we feel that our derivation of $\frac{c}{\underline{v}}$ which culminated in (9) is very satisfying, especially since it shows that we have made necessary headway for defining

*Again, while our present illustration is in terms of arrows, notice that $\|\underline{v}\|$ has been defined for all vector spaces. In keeping with the spirit of the metric used in the "arrow" cases, we do agree to use the Euclidean metric rather than the Minkowski metric when we talk about $\frac{\underline{v}}{\|\underline{v}\|}$. Moreover, as we saw in Unit 1 of this Block, if we want to use the general result that $\underline{x}\cdot\underline{y} \leq \|\underline{x}\|\ \|\underline{y}\|$ we must use the Euclidean metric since the result is not true for the Minkowski metric.

$$\lim_{\Delta \underline{x} \to \underline{0}} \frac{f(\underline{a}+\Delta \underline{x})-f(\underline{a})}{\Delta \underline{x}} .$$

Yet there is something about (9) which tends to make the mathematician cautious. Specifically, we would most likely want to be able to think of $\frac{c}{\underline{v}}$ as a single vector rather than as an infinite set of vectors (in much the same way that we need not distinguish between the number $\frac{a}{b}$ and the set $\frac{a}{b}$ when $b \neq 0$, except that then the choice was easy since the set $\frac{a}{b}$ has only a single element).

But how shall we go about the process of selecting a single, well-defined member of the set described by (9)? Well, the one thing that each member of the set shares in common is that its projection in the direction of $\underline{v}$ is $\frac{c}{\|\underline{v}\|}$*. If this is the only property that is of interest to us, why not choose that member of $\frac{c}{\underline{v}}$ which already has the direction of $\underline{v}$?** After all, in the expression $\frac{c}{\underline{v}}$, $\underline{v}$ is the only vector which is specifically named.

---

*While, admittedly, it is easier to think in terms of arrows than to think abstractly, keep in mind from our discussion in Chapter 4 of these notes that "direction" is defined for all spaces. Namely, "$\underline{x}$ has the same direction as $\underline{v}$" simply means that $\underline{x}$ is a scalar multiple of $\underline{v}$, or, in n-tuple notation, given $\underline{v} = (v_1,\ldots,v_n)$, then the set of all vectors which have the same direction as $\underline{v}$ is defined to be $\{t\underline{v}: t \text{ a real number}\}$, i.e., $\{(tv_1,\ldots,tv_n): t \text{ any real number}\}$.

**While $\frac{c}{\underline{v}}$ has infinitely many members (if $\underline{v} \neq \underline{0}$) the membership is still rather selective. Namely, if we insist on a particular direction and sense there is one and only one $\underline{x} \in \frac{c}{\underline{v}}$ such that $\underline{x}\cdot\underline{v} = c$. Pictorially, this is easy to see, for although infinitely many vectors have the same projection in a given direction, no two vectors whose directions (and sense) are equal but whose magnitudes are unequal can have the same projection. More analytically, in terms of a specific example

$$\frac{78}{(3,1,2,3,4)} = \{(x_1,x_2,x_3,x_4,x_5): 3x_1+x_2+2x_3+3x_4+4x_5=78\}$$

is an infinite set. If we now select all members of the set which have the same direction as (3,1,2,3,4) we mean that our members have the form (3t,t,2t,3t,4t), and hence

$3(3t)+(t)+2(2t)+3(3t)+4(4t) = 78$; or $t = 2$

Therefore, (6,2,4,6,8) is the only member of $\frac{78}{(3,1,2,3,4)}$ which has the same direction (and sense) as (3,1,2,3,4).

Since $\frac{\underline{v}}{\|\underline{v}\|}$ is the unit vector in the direction of $\underline{v}$, we have that the vector $\frac{c}{\underline{v}}$ would be written as

$$\frac{c}{\|\underline{v}\|} \frac{\underline{v}}{\|\underline{v}\|} \text{ or } \frac{c}{\|\underline{v}\|^2} \underline{v} \qquad (10)$$

With this as motivation, we now define $\frac{c}{\underline{v}}$ by

$$\frac{c}{\underline{v}} = \frac{c}{\|\underline{v}\|} \frac{\underline{v}}{\|\underline{v}\|} = \frac{c}{\|\underline{v}\|^2} \underline{v} \qquad (11)$$

and our previous discussion insures that by (11), $\underline{v} \cdot \frac{c}{\underline{v}} = c$.

As a computational review, we may compute $\underline{v} \cdot \frac{c}{\underline{v}}$ using the fact that $\underline{v} \cdot \underline{v} = \|\underline{v}\|^2$*, so that $\underline{v} \cdot \frac{c}{\|\underline{v}\|^2} \underline{v} = \frac{c}{\|\underline{v}\|^2} \|\underline{v}\|^2 = c$.

C

The Directional Derivative in n-Dimensions

Lest we lose sight of the forest because of the trees, let us summarize the main computational points of the previous section. In an attempt to define $f'(\underline{a})$ by mimicking the definition of $f'(a)$ we are forced to

*Here we see a strong reason why we are using the Euclidean metric. Namely, if $\underline{v} = (v_1, \ldots, v_n)$ then $\underline{v} \cdot \underline{v} = v_1^2 + \ldots + v_n^2$. Now $\|\underline{v}\| = \sqrt{v_1^2 + \ldots + v_n^2}$ only if we are using the Euclidean metric. Hence for the Euclidean metric $\underline{v} \cdot \underline{v} = \|\underline{v}\|^2$. (For the Minkowski metric, $\underline{v} \cdot \underline{v}$ would still be $v_1^2 + \ldots + v_n^2$ since the dot product is defined without reference to any metric. However, now, $\|\underline{v}\|$ would be $\max\{|v_1|, \ldots |v_n|\}$ hence $\|\underline{v}\|^2 = \max\{v_1^2, \ldots, v_n^2\}$. Certainly, there is no reason to expect that

$$\max\{v_1^2, \ldots, v_n^2\} = v_1^2 + \ldots + v_n^2$$

In other words, $\underline{v} \cdot \underline{v} = \|\underline{v}\|^2$ need not be true for the Minkowski metric.)

"invent" a definition of what it meant to divide a number by a vector. After much experimentation in terms of the logical consequences, we accepted as the main definition:

If c is any real number and $\underline{v}$ is any non-zero vector then $\frac{c}{\underline{v}}$ is that vector which is in the direction of $\underline{v}$ and whose magnitude is $\frac{c}{\|\underline{v}\|}$.*

If we now return to the definition of $f'(\underline{a})$ as given by (2), we find that we have solved one problem but have created another. That is, we became so worried about what it meant to divide a number by a vector that we went no further with (2) but rather began an immediate investigation into how this quotient should be defined. Having solved this problem, we are now ready to discover that a new major problem looms before us. Namely, we have previously agreed (and for good reasons) that when we wrote $\lim_{\Delta\underline{x}\to\underline{0}}$ the meaning was that not only should the limit exist but its value must not depend on the direction by which $\Delta\underline{x}$ approached $\underline{0}$. On the other hand, one consequence of our definition of the quotient of a number divided by a vector is that the vector which is denoted by the quotient changes as $\Delta\underline{x}$ changes direction! That is, the quotient was defined as a vector in the direction of $\Delta\underline{x}$, so that, if $\Delta\underline{x}$ is not rigidly specified, the quotient is, in a sense, undefined since we have no way of determining the direction of the quotient.

Thus, it appears that another refinement is required before we can work with (2), and it is this refinement that motivates the meaning of a directional derivative. More specifically, it will happen quite in general that the limit in (2) will depend on the direction by which $\Delta\underline{x}$ approaches $\underline{0}$, for not only does changing the direction of $\Delta\underline{x}$ affect the denominator of our quotient, it affects the numerator as well since $f(\underline{a} + \Delta\underline{x})$ will, in general, depend on the direction of $\Delta\underline{x}$. Thus, it would appear that if we held to (2) without some modification, $f'(\underline{a})$ would never exist since the limit which defines $f'(\underline{a})$ will not

---

*Actually, it is possible that c is negative in which case $\frac{c}{\|\underline{v}\|}$ cannot be a magnitude (since magnitudes are non-negative). What we should say is that the magnitude is $\frac{|c|}{\|\underline{v}\|}$ and that if c is negative $\frac{c}{\underline{v}}$ has the opposite sense of $\underline{v}$.

exist (since for the limit to exist its value must not depend upon how $\Delta \underline{x} \to \underline{0}$). We shall try to clarify this point in a few moments by means of a specific example (with other examples being supplied in the Exercises), but first we prefer to remove the new pitfall.

We now agree to remove any ambiguity from (2) by specifying a particular direction. (It is often conventional to specify a direction in n-dimensional space in terms of a unit vector. Quite often, in fact, a unit vector is called a direction. For example, in terms of a planar example, $\frac{3}{5}\vec{i} + \frac{4}{5}\vec{j}$ is a unit vector in the direction of $3\vec{i} + 4\vec{j}$. In the modern vernacular, we would say $\frac{3}{5}\vec{i} + \frac{4}{5}\vec{j}$ is a direction, and any scalar multiple of this vector would be said to have the same direction.) In any event, if we let $\underline{u}$ denote a specific direction, we may think of $\Delta \underline{x}$ as always being in the direction of $\underline{u}$, and that in this context $\Delta \underline{x} \to \underline{0}$ means that the direction of $\Delta \underline{x}$ stays fixed but its magnitude approaches 0.

The next question is that of finding a way to change the notation in (2) to reflect this idea. To this end, we observe that once the direction $\underline{u}$ is fixed, all other vectors in this direction are of the form $t\underline{u}$ where t is a real number. Thus, rather than write $\Delta \underline{x}$, which carries the connotation of a varying direction, we write $t\underline{u}$. With this in mind, the left side of (2) becomes ambiguous since it does not indicate the direction $\underline{u}$. For this reason, we agree to rewrite $f'(\underline{a})$ as $f_{\underline{u}}'(\underline{a})$ and we call this the derivative of f at $\underline{a}$ in the direction $\underline{u}$.

At the same time, replacing $\Delta \underline{x}$ by $t\underline{u}$ converts the right side of (2) into

$$\lim_{t\underline{u} \to \underline{0}} \left[\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{t\underline{u}} \right] \tag{12}$$

and since $\underline{u}$ is a unit vector, $t\underline{u} \to \underline{0}$ if and only if $t \to 0$. Moreover, since $t\underline{u}$ and $\underline{u}$ are to have the same sense, $t \geqslant 0$; hence, $t \to 0$ may be replaced by $t \to 0^+$. Thus, (12) may be rewritten as

$$\lim_{t \to 0^+} \left[\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{t\underline{u}} \right] \tag{13}$$

The bracketed expression in (13) denotes the vector in the direction of ($\underline{u}$) whose magnitude is

$$\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{\|t\underline{u}\|}$$

and this in turn is

$$\left[\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{\|t\underline{u}\|}\right]\underline{u} \quad .$$

Then since $\|t\underline{u}\| = |t|\ \|\underline{u}\|$ (and since $t>0$, $|t| = t$) and $\|\underline{u}\| = 1$ by virtue of being a unit vector, we have

$$\left[\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{\|t\underline{u}\|}\right]\underline{u} = \left[\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{t}\right]\underline{u} \qquad (14)$$

The bracketed expression on the right side of (14) represents the average rate of change of $f(\underline{x})$ in the direction $\underline{u}$ with respect to $\|\underline{x}\|$ as $\underline{x}$ varies from $\underline{x} = \underline{a}$ to $\underline{x} = \underline{a}+t\underline{u}$. Thus

$$\lim_{t\to 0^+} \frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{t\underline{u}}$$

represents the instantaneous rate of change of $f(\underline{x})$ with respect to $\|\underline{x}\|$ in the direction $\underline{u}$ at $\underline{x} = \underline{a}$.

If we make these changes (2) becomes

$$f'_{\underline{u}}(\underline{a}) = \left\{ \lim_{t\to 0^+} \left[\frac{f(\underline{a}+t\underline{u}) - f(\underline{a})}{t}\right]\right\} \underline{u} \qquad (15)$$

Equation (15) summarizes rather nicely the idea that $f'_{\underline{u}}(\underline{a})$ is a vector in the direction (of) $\underline{u}$ whose magnitude is the derivative of $f(\underline{x})$ with respect to $\|\underline{x}\|$ at $\underline{x} = \underline{a}$ in the direction $\underline{u}$.

Thus, it is natural to view $f'_u(\underline{a})$ as a directional derivative, that is, a derivative in the direction $\underline{u}$. Notice that if we use the generalized definition of direction that applies to all dimensional vector spaces, then (15) is valid for any dimensional space, and, in particular, if we restrict our attention to the special case of planar arrows, it should be easy to see that (15) is equivalent to the definition of a directional derivative as given in the text and studied as part of the previous unit.

Perhaps a specific example in which we compute a directional derivative in the plane using each of the two methods and then show that the answers are the same, might be of more benefit than an attempt to make an abstract, formal proof. To this end, let us consider the problem of finding the directional derivative of the surface $w = x^2y$ at the point (2,3,12) in the direciton of the vector $3\vec{i} + 4\vec{j}$. Using the traditional approach, we have that $f(\underline{x}) = f(x,y) = x^2y$, whence $f_x(x,y) = 2xy$ and $f_y(x,y) = x^2$. Therefore, $f_x(2,3) = 12$ and $f_y(2,3) = 4$. Accordingly, the gradient of f at (2,3) is $12\vec{i} + 4\vec{j}$, and a unit vector in the given direction is $\frac{3}{5}\vec{i} + \frac{4}{5}\vec{j}$. Since the desired directional derivative is the dot product of the gradient and the given unit vector, we obtain as the directional derivative

$(12\vec{i} + 4\vec{j}) \cdot (\frac{3}{5}\vec{i} + \frac{4}{5}\vec{j})$, or

$$\frac{36}{5} + \frac{16}{5} = \frac{52}{5}$$

If we now use the method of this section, we have

$$f(\underline{x}) = f(x,y) = x^2y$$

$$\underline{a} = (2,3)$$

$$\underline{u} = (\tfrac{3}{5}, \tfrac{4}{5})$$

$$t\underline{u} = (\tfrac{3t}{5}, \tfrac{4t}{5})$$

Therefore,

$$\underline{a} + t\underline{u} = (2 + 3t/5,\ 3 + 4t/5) = \left(\frac{10+3t}{5},\ \frac{15+4t}{5}\right)$$

and

$$f(\underline{a} + t\underline{u}) = \left(\frac{10+3t}{5}\right)^2 \left(\frac{15+4t}{5}\right)$$

$$= \frac{1500 + 1300t + 375\ t^2 + 36\ t^3}{125}$$

$$= 12 + 52t/5 + 3\ t^2 + 36\ t^3/125.$$

Then, since $f(\underline{a}) = f(2,3) = 12$ we have

$$f(\underline{a} + t\underline{u}) - f(\underline{a}) = 52t/5 + 3t^2 + 36t^3/125, \text{ and}$$

$$\frac{f(a + tu) - f(a)}{t} = 52/5 + 3t + 36t^2/125.$$

If we now let $t \to 0^+$, we obtain from (15)

$$f'_{\underline{u}}(\underline{a}) = (52/5)\underline{u} \qquad (16)$$

which is a vector whose magnitude is 52/5 and whose direction is $\underline{u} = \frac{3}{5}\vec{i} + \frac{4}{5}\vec{j}$, and this, of course, has the same direction as $3\vec{i} + 4\vec{j}$. We thus see that we obtain the same answer in both cases.

Let us also observe that the method of this chapter does not require that we be aware of the concept of the gradient (that is, in deriving (16) we never used anything but the expressions implied in (15), and certainly no notion of the gradient is present there). Not only did we not need the gradient, but we had no need to talk about partial derivatives. This is as it should be, since the partial derivatives are merely derivatives with respect to some highly selective directions. In this respect we can compute $f'_{\underline{u}}(\underline{a})$ from (15) in any vector space, without regard to either a gradient or partial derivatives. Such additional examples are left for the exercises. Before ending this section, however, we feel it might provide a new insight to calculus of a single real variable if we apply the discussion of this section to a 1-dimensional vector space.

To this end, observe that in 1-dimensional space our vectors (1-tuples) are simply real numbers. Thus, the vector $\underline{a}$ may be identified with the number, a. Moreover, in 1-dimensional space there are only the two unit vectors, $\vec{i}$ and $-\vec{i}$. (Geometrically, the space is the x-axis and along the x-axis any vector is a scalar multiple of $\vec{i}$.)

Now, recall that one way of saying that f'(a) existed was to say that both

$$\lim_{\Delta x \to 0^+} \left[\frac{f(a+\Delta x) - f(a)}{\Delta x} \right] \quad \text{and} \quad \lim_{\Delta x \to 0^-} \left[\frac{f(a+\Delta x) - f(a)}{\Delta x} \right]$$

exist <u>and</u> are equal.

But

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

is a directional derivative. It is the derivative in the direction that $\Delta x \to 0$ through positive values, that is, from right-to-left, and this, in turn, is the direction $-\vec{i}$. In other words, using the notation of this section,

$$\left.\begin{array}{l} \lim_{\Delta x \to 0^+} \left[\frac{f(a+\Delta x) - f(a)}{\Delta x} \right] = f'_{-\vec{i}}(a) \\ \text{and} \\ \lim_{\Delta x \to 0^-} \left[\frac{f(a+\Delta x) - f(a)}{\Delta x} \right] = f'_{\vec{i}}(a) \end{array}\right\} \qquad (17)$$

Thus, from (17) our new language says that $f'(\underline{a})$ exists in 1-dimensional space if and only if the two directional derivatives (where both directions differ only in sense) exist and are equal.

D

The Derivative in n-Space

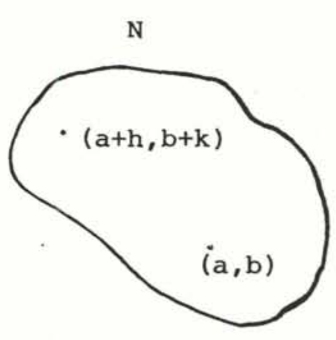
Our attempts to define the derivative of a function $f:E^n \to E$ by mimicking the 1-dimensional case has led to the "invention" of the directional derivative. The trouble with the directional derivative is that it may exist in some directions but not in others. This is easy to picture in the case $f:E^2 \to E$ since then the graph of f is a surface, and what we are then saying is that some cross sections of this surface through a given point may be smooth while others aren't.

If we still think of the derivative as denoting "smoothness" (without worrying about what this means in high dimensions), then unless the directional derivative exists in every direction (i.e., every slice through the point is smooth) the surface will not be smooth. Stated more positively, we define a surface to be smooth at a point if the directional derivative exists in each direction at the given point.

With this as motivation, it is an easy step to generalize the definition (even if the geometric interpretation may no longer apply). Namely, we say that $f:E^n \to E$ is differentiable at $\underline{x} = \underline{a}$ if and only if $f'_{\underline{u}}(\underline{a})$ exists in every direction, $\underline{u}$.

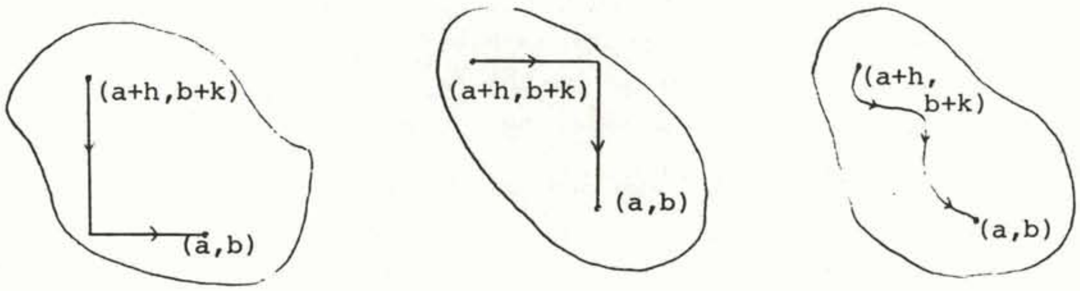
While the above discussion is an adequate intuitive motivation, we can support the argument for our definition of differentiability on more computational grounds as well. Namely, we would like the definition of a derivative to be independent of any particular choice of direction. Why is this so? Perhaps the following pictorial demonstration in the case n=2 will shed some light on the answer.

Suppose f(x,y) is defined in some neighborhood N of (a,b) and we choose (c,d) to be any other point in this neighborhood. To emphasize that (a,b) is our focal point, we shall rewrite (c,d) as (a+h, b+k) where h and k are constants which depend on the choice of the point (c,d). [To be more specific, h is what we would ordinarily call Δx and k is Δy].



(Figure 3)

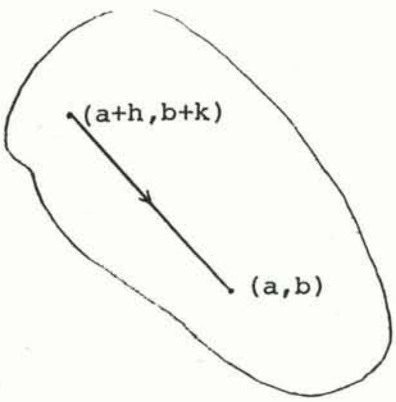
Somehow we expect that the derivative of f, no matter how we ultimately define it, must involve Δf [which denotes f(a+h, b+k) - f(a,b)]. Now, once h and k are specified, f(a+h, b+k) - f(a,b) is a well-defined number which in no way depends on direction. Moreover, when we finally take the appropriate limit, we want an answer which will not depend on the path which joins (a+h, b+k) to (a,b); for if the answer does depend on the path, the limit does not exist. To be sure, the usage of $f_x(a,b)$ and $f_y(a,b)$ utilize paths such as in Figures 4a and 4b, but our answer must hold for arbitrary (and not necessarily straight line) paths as in Figure 4c.



(a) (b) (c)

(Figure 4)

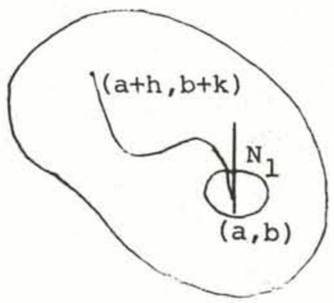
The special role of the directional derivative in this context is shown in Figure 5.



(Figure 5)

That is, once (a+h,b+k) is chosen the directional derivative assumes that the path is the straight line which joins (a+h,b+k) to (a,b). In this sense, once the straight line is determined, the points (x,y) for which we evaluate f(x,y) all lie on this line, so that in computing a directional derivative we do have the analog of the 1-dimensional derivative. The only problem is that this derivative, as we have seen, varies with direction.

We should also point out that while, as in Figure 4, our paths do not have to be straight lines, we may assume (at least intuitively in the case n=2) that they are straight lines. Namely, as we take limits, we are only interested in what happens "near" the point (a,b). Thus, if near (a,b) the path which joins (a+h,b+k) to (a,b) is not a straight line, we may replace it by the straight line which is tangent to the path at (a,b), assuming of course, that the path is smooth. (See Figure 6)



In N_1 the path and the tangent line are essentially the same.

(Figure 6)

Thus, this discussion, too, motivates why in formulating the definition of the derivative of a function of several (n) real variables we first insist that the directional derivative of f at $\underline{a}$ exist in every direction.

Now that we have given a few motivations for defining f to be differentiable if its directional derivative exists in each direction, the next subtle point is the choice of how the derivative of f should be defined. That is, there is a difference between saying that f is differentiable and determining what the derivative actually is. For reasons that we hope will be made clearer as we proceed, let us define the derivative of f at $\underline{x} = \underline{a}$ to be the directional derivative of f at $\underline{a}$ which has the greatest magnitude. Notice that this introduces the additional subtlety that there must be a derivative of maximum magnitude. This is not at all self-evident (and indeed it need not be true, although we shall not pursue this matter here) since there are infinitely many directional derivatives of f at $\underline{a}$, and for an infinite set there need not be a (finite) upper bound.

While it may not be clear yet why we choose this definition, what should be clear is that since the directional derivative varies from direction to direction, we must somehow or other make a choice that picks one of these values from all others (in much the same way as we had to choose one vector from the set of vectors, $\frac{\underline{c}}{\underline{v}}$). How shall we make this choice? As usual, we shall let the most important application of the concept determine the definition. In this case, we find that we shall usually be trying, in one form or another, to make the expression $|f(\underline{x}) - f(\underline{a})|$ "sufficiently small" regardless of how $\underline{x}$ approaches $\underline{a}$, but if we want $|f(\underline{x}) - f(\underline{a})|$ to be sufficiently small, then we need only insure that it is sufficiently small in the direction in which it is the greatest.

If we now let $f'(\underline{a})$ denote the directional derivative of f at $\underline{a}$ whose magnitude is maximum, this problem is taken care of. (A more formal way of saying this is that the directional derivative behaves like an ordinary 1-dimensional derivative in the given direction. That is, if $f'_{\underline{u}}(\underline{a})$ exists then

$$\Delta f = f'_{\underline{u}}(\underline{a}) \cdot \Delta\underline{x} + \underline{k} \cdot \Delta\underline{x},$$

where $\lim_{\Delta\underline{x}\to 0} \underline{k} = \underline{0}$ and $\Delta\underline{x}$ is in the direction, $\underline{u}$. All we are then saying is that Δf is maximum when $f'_{\underline{u}}(\underline{a})$ is maximum in magnitude.)

At any rate, this completes our supplementary discussion of the derivative of a function of several variables. As a review, so that we see where all the pieces fit into place, let us observe that in Unit 3 we discussed the gradient and the directional derivative in terms of f_x and f_y (using the 2-dimensional notation). While it was not specifically mentioned in the text, the concept of differentiability of f at $\underline{x} = \underline{a}$ required that f_x and f_y exist at $\underline{a}$ and be continuous there as well.

While this was a very good approach, the point was that the use of f_x and f_y as the basic building block was a departure from our attempts to show all results as extensions of the 1-dimensional case. Thus, it was not so much that we produced different results in this section (in fact, we didn't) but that we were able to re-derive the results of the text from the point of view of our emphasis on structure.