A

Introduction

By way of a brief review of some ideas introduced in Chapter 2 and 3 of these notes, recall that once we agree that our variables may be either scalars (numbers) or vectors, the traditional notation, $f(x)$, now has four interpretations. They are:

(1) $f(x)$

(2) $\vec{f}(x)$

(3) $f(\vec{x})$

(4) $\vec{f}(\vec{x})$

Case (1) was handled as Part 1 of this course. Namely, in (1) both our "input" and "output" are real numbers, and this is precisely what is meant by a real-valued function of a single real variable.*

Case (2) was handled in Block 2 of this course. To be sure, we particularly emphasized the special forms

$\vec{R}(t)$, $\vec{v}(t)$, and $\vec{a}(t)$

but the point is that we were studying vector functions of a single real variable.

Case (4) will be discussed in a different context later in this block. It is worth pointing out that when we begin our study of Complex Variables (Part 3) we will be studying a special case of (4). That is, from a geometrical point of view, it is conventional to view a

*Frequently, this "mouthful" is abbreviated as "function of a single variable." Unless the meaning is clear from context, this is a very ambiguous description. For example, in the last block when we wrote such expressions as $\vec{R} = \vec{f}(t)$, we meant that $\vec{R}$ was also a function of a single variable, but $\vec{R}$ now was a vector function (i.e., the output was a vector). Moreover, in an example where we might be studying temperature, T, as a function of position in space, $\vec{R}$, we have $T = f(\vec{R})$, which again indicates a function of a single variable. Yet, in this case, the function is a scalar while the "input" is a vector. That is, we have in this case a real-valued function of a single vector variable. In particular, each of the cases (1), (2), (3), and (4) above are examples of "functions of a single variable."

complex number as a vector in the plane (since we may think of the complex number as having two components, one called the real part and the other called the imaginary part). Thus, if both the input and the output of our function machine are complex numbers (in more formal language, if we have a complex-valued function of a single complex variable) then we may view this situation as a vector function of a single vector variable, which is precisely what is being described in case (4).

Our aim in this chapter is to study case (3) in detail. In a sense, this study is an "inversion" of the study in Chapter 3, since in that chapter the input was a scalar and the output was a vector, while here it is the output which is the scalar and the input which is a vector.

B

Scalar Functions of Vector Variables

Certainly, there are a multitude of physical examples in which we measure a scalar variable in terms of a vector variable. We might be considering work done on a particle in terms of the force exerted on it [$w = w(\vec{F})$], or we might be considering the temperature of a particle in terms of its position in space [$T = T(\vec{R})$].

For the purpose of our present discussion, let us, for simplicity, restrict our study to the case where our vectors are 2-dimensional. Suppose we have a temperature distribution given by the (unlikely) formula

$$T(\vec{R}) = xy, \text{ where } \vec{R} = x\,\vec{i} + y\,\vec{j}. \tag{1}$$

Substituting for $\vec{R}$ its description in terms of i and j components, (1) becomes

$$T(x\,\vec{i} + y\,\vec{j}) = xy. \tag{2}$$

If we now introduce our abbreviation that (x,y) denotes $x\,\vec{i} + y\,\vec{j}$, then (2) becomes

$$T\big((x,y)\big) = xy. \tag{3}$$

Since the notation $T\big((x,y)\big)$ is cumbersome, it is a generally agreed-upon convention to abbreviate it by the simpler notation, $T(x,y)$.

[This statement does not depend on T. That is, if f is any scalar function of the vector (x,y), it is conventional to write $f(x,y)$ rather than $f\big((x,y)\big)$.]

In any event, using this convention, (3) takes the form

$$T(x,y) = xy. \tag{4}$$

Pictorially, (4) may be represented by

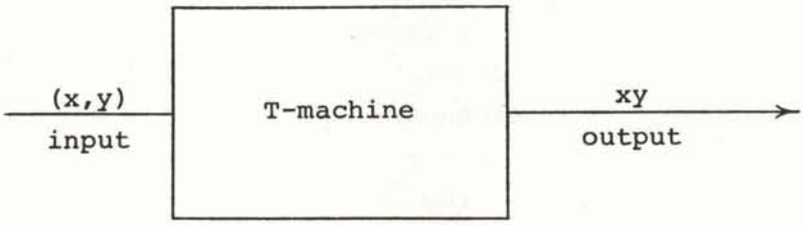


Figure 1

We now come to the most crucial point of this block. If we look at (4) without knowing how it was derived, it would appear that T was a real function not of a vector but of <u>two real variables</u>.

For example (and this is why we chose such a far-fetched temperature distribution), suppose x were to denote the length of the base of a rectangle, y the height, and T the area of the rectangle. Then in the language of traditional mathematics, we would write

$$T = T(x,y) = xy. \tag{5}$$

The point is that without knowing the context, we <u>cannot</u> distinguish between (4) and (5). In other words, if we were to draw a function machine illustration for the case of the area of a rectangle being the product of its base and height, the diagram would look identical to Figure 1!

Yet, there is something about the context of the variable in the example that led to (5) that makes us uneasy thinking about (x,y) as a vector, at least in the sense of our tendency to identify vectors with arrows. In still other words, we do not tend to think of the dimensions of a rectangle as being the components of an "arrow." Yet (x,y) is as bona fide a 2-tuple (see the introduction to Chapter 2 for a review of the n-tuple notation) for denoting the dimensions of a rectangle as it is for denoting the $\vec{i}$ and $\vec{j}$ components of a planar vector. The only difference is that in one case it is easy to think in terms of arrows while in the other case it isn't.

For this reason, we prefer to think in terms of 2-tuples rather than arrows in the plane. That is, the 2-tuple idea makes sense regardless of whether it is "natural" to think in terms of arrows. At the same time, if the 2-tuple does lend itself to an arrow interpretation, we simply have an additional way of viewing the 2-tuple.

If we now agree to replace the notion of a planar vector (arrow) by that of a 2-tuple, we see that every real function of two real variables is a special case of $f(\vec{x})$ provided we rewrite $f(\vec{x})$ as $f(\underline{x})$ where, in deference to our remark that a 2-tuple makes sense without the need for an arrow interpretation, we have switched from the notation $\vec{x}$ to the notation $\underline{x}$. That is, it is our feeling that $\underline{x}$ is a more neutral notation than $\vec{x}$ if we wish to "play down" the arrow interpretation.

By way of an example, suppose we let $\underline{x}$ denote (x_1, x_2) and we are given the equation

$$y = f(\underline{x}) = x_1^{\,3} + 4x_2. \tag{6}$$

Then we certainly have the right to interpret (6) in terms of arrows by saying that f maps the vector $x_1\vec{i} + x_2\vec{j}$ into the scalar $x_1^{\,3} + 4x_2$. In this context, for instance, we would have

$$f(5\,\vec{i} + 6\,\vec{j}) = (5)^3 + 4(6) = 149.$$

On the other hand, we have equal right to interpret equation (6) in the traditional way that y is a function of the two real variables x_1 and x_2, and in particular when $x_1 = 5$ and $x_2 = 6$, $y = (5)^3 + 4(6) = 149$.

Notice that while we do not advocate which of the two interpretations of (6) is the better, we hope that it is clear that the second interpretation includes the first as a special case, but that the first interpretation seems too "specialized" to include the second. We do not intend to pursue this notion further here. Rather, we only want to establish the important point that it is advantageous to start thinking of planar vectors as 2-tuples rather than as arrows (and, in this regard, to start thinking of spatial vectors as 3-tuples rather than as arrows). This idea is far too important to be brushed off lightly, but further discussion of it at this time is a digression from our main theme. For this reason, additional discussion is left to the last section of this chapter.

While we feel that the preceding discussion was motivation enough for making the notational change from $\vec{x}$ to $\underline{x}$, the next major point is that we have not yet scratched the surface in explaining the real significance of this change in notation. In this respect, what we would next like to point out is the following. It should seem clear that our usual experience with geometry makes it rather self-evident that the quantity named by $\vec{x}$ denotes either a 1-tuple (the vector is parallel to the x-axis, for example), a 2-tuple (the vector lies in the xy-plane), or a 3-tuple (the vector lies in xyz-space). We certainly would not be tempted to think beyond 3-tuples, if only because we view an arrow as a geometric entity, and, as such, it has no meaning beyond 3-dimensional space.

On the other hand, as mentioned in the introduction to Chapter 2, the notion of an n-tuple is perfectly well defined (i.e. very meaningful) even if n exceeds 3. Returning to our study of temperature distributions, for example, if we are interested in studying temperature distribution in the room, we usually find that the temperature depends on where we are in the room (which in Cartesian coordinates means that we must know the x, y, and z components of the point at which we are making the measurement in the room) and, at a fixed location in the room, usually also depends on time (i.e. in general, the thermometer registers different readings at different times even though it stays in the same position).

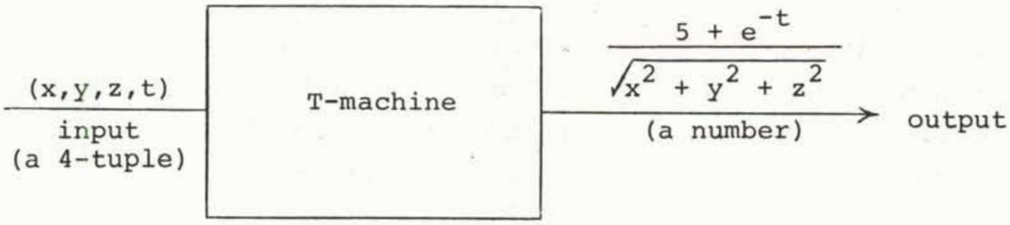
The most general way of representing this idea in the language of functions is to write

$$T = T(x,y,z,t). \tag{7}$$

To make (7) seem more concrete, we might have the situation where the room is centrally heated in such a way that the temperature is inversely proportional to the distance from the heat source to a point in the room and perhaps as time goes on the temperature tends to "level off." For example, we might have (and again we are making no attempt to conform to reality as the physicist sees it) as an example of (7),

$$T = T(x,y,z,t) = \frac{5 + e^{-t}}{\sqrt{x^2 + y^2 + z^2}}. \tag{8}$$

In terms of a function machine, (8) may be viewed as



In other words, with respect to the present example, while a notation like $\vec{x}$ might seem unnatural for denoting the 4-tuple (x,y,z,t), there is nothing unnatural about letting $\underline{x}$ denote (x,y,z,t). More generally, there is no need to think specifically in terms of x, y, z, and t. Rather, we may let x_1, x_2, x_3, and x_4 denote any four real variables, whereupon we may then let

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

In this way, we may in a meaningful way abbreviate

$$f(x_1, x_2, x_3, x_4)$$

by

$$f(\underline{x})$$

and with the latter notation, functions of several real variables begin to resemble the form of functions of a single real variable.

In other words, looking at an expression such as $f(\underline{x})$, we are tempted to mimic certain definitions that were used in our study of real functions of a single real variable. For example, we might be tempted to extend such concepts as $\lim_{x \to a} f(x)$ to $\lim_{\underline{x} \to \underline{a}} f(\underline{x})$. Intuitively, we probably would feel, just as in the scalar case, that

$$\lim_{\underline{x} \to \underline{a}} f(\underline{x}) = L$$

means that $f(\underline{x})$ can be made as arbitrarily nearly equal to L as we wish simply by picking $\underline{x}$ "sufficiently close" to $\underline{a}$. (Notice that we write L not $\underline{L}$ since $f(\underline{x})$ is a scalar, not a vector.)

Notice that a rather interesting "confrontation" now takes place. Certainly, we have in the previous block talked about what it meant for two vectors (arrows) to be "near" each other. For example, this occurred when we discussed such things as $|\vec{R}(t + \Delta t) - \vec{R}(t)| < \varepsilon$. The point is that, in these cases, we had a physical (geometrical) meaning for the difference of two vectors in terms of our concept of addition whereby we placed the arrows head-to-tail, etc. Clearly, such an interpretation presupposes that we were restricted to no more than spatial vectors, for, indeed, the concept of arrows with heads and tails does not extend beyond the study of 3-dimensions.

Thus, in a sense, we open Pandora's Box when we now allow $\underline{x}$ and $\underline{a}$ to denote n-tuples, for in this case, how shall we replace the notion of adding vectors (n-tuples) by placing them head-to-tail? This shall be the topic of our next section, after which we shall return to a further study of functions of the form $f(\underline{x})$ where $\underline{x}$ denotes an n-tuple.

C

An Introduction to n-tuple Arithmetic

There is a saying in mathematics that objects are known by the company they keep. What we mean by this is that mathematics is interested in structures rather than in merely sets of objects. For example, in elementary arithmetic, one first learns to count. This does not become arithmetic until we define such things as equality, addition, etc. In a similar vein, when we introduced vectors, we talked about arrows. We did not talk about vector arithmetic until we first defined what it meant for two arrows to be equal, how we were to add two arrows, and how we were to multiply an arrow by a scalar.

Once again, we are at this crossroads. That is, we have now defined what we mean by an n-tuple. Yet, from a structural point of view, we are powerless since we have no way, as yet, to perform an arithmetic of n-tuples. For example, what shall it mean for two n-tuples to be equal? Or, how shall we add two n-tuples?

The point is that we learned how to do this, even though we may not have realized it, when we were studying "arrows" - i.e., when we were studying n-tuples with n = 1, 2, or 3.

Namely, when we converted our "arrow" definitions into Cartesian coordinates, and used the n-tuple notation, we saw that

$(a_1,a_2,a_3) = (b_1,b_2,b_3)$ means that (9) meant that

$a_1 = b_1,\ a_2 = b_2,$ and $a_3 = b_3$

$$(a_1,a_2,a_3) + (b_1,b_2,b_3) = (a_1 + b_1,\ a_2 + b_2,\ a_3 + b_3) \tag{10}$$

and

$$c(a_1,a_2,a_3) = (ca_1,ca_2,ca_3) \tag{11}$$

where we have elected to write (9), (10) and (11) in terms of spatial vectors, noting that similar results hold for 2-tuples and 1-tuples.

Now, while it is true that (9), (10) and (11) were motivated by the "arrow" interpretation, the fact remains that once (9), (10) and (11) are stated they make sense in their own right without any reference to arrows.

We now elect to use (9), (10) and (11) to define an arithmetic on any set of n-tuples. Since the case n = 4 is probably alien enough to you, we shall use n = 4 rather than a general n so that you may get a concrete idea of what is happening here.

To begin with, we have the set S, say, of all 4-tuples. That is, $S = \{(x_1,x_2,x_3,x_4): x_1,\ldots,x_4$ are real numbers$\}$. Letting $\underline{a} = (a_1,a_2,a_3,a_4)$ and $\underline{b} = (b_1,b_2,b_3,b_4)$ denote arbitrary members of S, we define an equivalence relation (=) on S by

$\underline{a} = \underline{b}$ means $a_1 = b_1,\ a_2 = b_2,\ a_3 = b_3,$ and $a_4 = b_4$.

(It is left as an exercise to verify that "=" defined in this way is indeed an equivalence relation on S. The fact that it is follows from the fact that "=" is an equivalence relation for the real numbers, but we will say more about this in the solution to the exercise.)

We then define $\underline{a} + \underline{b}$ to be the 4-tuple $(a_1 + b_1,\ a_2 + b_2,\ a_3 + b_3,\ a_4 + b_4)$. In other words, with $\underline{a} = (a_1,a_2,a_3,a_4)$ and $\underline{b} = (b_1,b_2,b_3,b_4)$, then $\underline{a} + \underline{b} = \underline{c}$ where $\underline{c} = (a_1 + b_1,\ a_2 + b_2,\ a_3 + b_3,\ a_4 + b_4)$.

Finally, if c denotes any scalar, we define $c\underline{a}$ by (ca_1,ca_2,ca_3,ca_4).

The key point now is that if the set S is endowed with the structure described above (i.e., with the given definition of "=", "+," and scalar multiplication) then and only then do we call the resulting structure a 4-dimensional vector space. In other words, it is the 4-tuples together with the above structure that is called the vector space, not the set of 4-tuples alone. This is analogous to our earlier remark that even in the arrow case, we do not talk about vector arithmetic until we have rules for combining and equating vectors. The arrows by themselves are not of too much use to us from a structural point of view.

Clearly, these results for $n = 4$ immediately generalize. In particular, our general definition of n-dimensional vector space is the following.

By the n-dimensional vector space E^n (to use the language of the text), we mean the set of all n-tuples together with the following structure:

(i) If $\underline{a} = (a_1,\ldots,a_n)$ and $\underline{b} = (b_1,\ldots,b_n)$ then $\underline{a} = \underline{b}$ means that $a_1 = b_1,\ldots,a_n = b_n$.

(ii) With $\underline{a}$ and $\underline{b}$ as above, $\underline{a} + \underline{b}$ is defined to be $\underline{c}$ where $\underline{c} = (a_1 + b_1,\ldots,a_n + b_n)$.

(iii) If c is any scalar, scalar multiplication is defined by $c(a_1,\ldots,a_n) = (ca_1,\ldots,ca_n)$.

What should now be noted is that everything we proved about arrows (with respect to equality, addition, and scalar multiplication) holds true for n-tuples (vectors) in n-dimensional space.

In particular, we see that $\underline{0} = (0,\ldots,0)$ plays the role of the additive identity since

$$\begin{aligned}
\underline{a} + \underline{0} &= (a_1,\ldots,a_n) + (0,\ldots,0) \\
&= (a_1 + 0,\ldots,a_n + 0) \quad \text{[by how "+" is defined in (ii)]} \\
&= (a_1,\ldots,a_n) \quad \text{[by the property of the real number, 0]} \\
&= \underline{a}.
\end{aligned}$$

Similarly, if we still elect to keep the structural property that $(-\underline{a})$ is defined by $\underline{a} + (-\underline{a}) = \underline{0}$, then we see almost immediately that if $\underline{a} = (a_1,\ldots,a_n)$ then $-\underline{a} = (-a_1,\ldots,-a_n)$, since

$$\underline{a} + (-\underline{a}) = (a_1,\ldots,a_n) + (-a_1,\ldots,-a_n)$$

$$= (a_1 - a_1,\ldots,a_n - a_n)^*$$

$$= (0,\ldots,0)$$

$$= \underline{0}.$$

It now makes sense to talk about the difference of two elements of E^n. That is, just as in the arrow cases, we may define $\underline{a} - \underline{b}$ to mean $\underline{a} + (-\underline{b})$. Our point is that we have now introduced all the necessary ingredients for generalizing the concept of "distance" to n-dimensional spaces, and once this is done, we shall have no trouble interpreting what we mean when we say, for example, that $\underline{x}$ is near $\underline{a}$, no matter what the dimension of the space.

How may we use our knowledge of vectors as arrows to determine the distance between $\underline{a}$ and $\underline{b}$ in any n-dimensional vector space? For reasons that we will try to make clear a bit later, let us denote the distance between $\underline{x}$ and $\underline{a}$ by $d(\underline{x},\underline{a})$. In terms of 1, 2, and 3-dimensional "arrow" spaces, the notation $d(\underline{x},\underline{a})$ was not used. Rather, we used $|\vec{x} - \vec{a}|$. (Actually, to mimic this notation we could have written $|\underline{x} - \underline{a}|$ rather than $d(\underline{x},\underline{y})$, but because of reasons of our own, we prefer not to use the absolute-value notation [thus, restricting the use of absolute values for numbers rather than vectors].)

By way of a very quick review, notice that in terms of the arrow interpretation and using $d(\underline{x},\underline{y})$ rather than $|\vec{x} - \vec{y}|$, we had that if $n = 1$

$$d(\underline{x},\underline{a}) = \sqrt{(x - a)^2}.$$

If $n = 2$

$$d(\underline{x},\underline{a}) = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2},$$

where $\underline{a} = (a_1,a_2)$ and $\underline{x} = (x_1,x_2)$. If $n = 3$

*Here we have taken the liberty of writing $a_1 - a_1$, etc. without actually exhibiting the intermediate step of first writing $a_1 + (-a_1)$, etc.

$$d(\underline{x},\underline{a}) = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2 + (x_3 - a_3)^2}$$

where $\underline{a} = (a_1,a_2,a_3)$ and $\underline{x} = (x_1,x_2,x_3)$.

With this as background, we now play our "usual" game and define $d(\underline{x},\underline{a})$ for any n-dimensional space by:

$$d(\underline{x},\underline{a}) = \sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2} \tag{12}$$

where $\underline{a} = (a_1,\dots a_n)$ and $\underline{x} = (x_1,\dots,x_n)$.

Notice that (12) captures the feeling of "closeness." For example, it seems rather intuitive that if we were told that $\underline{x} = (x_1,\dots,x_n)$ was close to $\underline{a} = (a_1,\dots,a_n)$ and we were supplied with no further hints, we would assume that it meant that x_1 was near a_1 (and here we know what "near" means since x_1 and a_1 are real numbers, and we studied this notion as Part 1 of our course), x_2 was near a_2, and..., x_n was near a_n.

The point is that (12) says this. Namely, the only way that a sum of squares can be small (since each square is non-negative) is for each of the numbers being squared to be small. Thus, from (12) we see that the only way $\underline{x}$ can be near $\underline{a}$ [i.e., $d(\underline{x},\underline{a})$ is small] is if each of the quantities $(x_1 - a_1),\dots,$ and $(x_n - a_n)$ is small in magnitude. 'his, in turn, is what we mean, when we say that x_1 is near $a_1,\dots,$ and x_n is near a_n.

Since the notation $d(\underline{x},\underline{a})$ may seem strange to you and since, on the other hand, we do not want to use absolute value symbols, let us compromise and from this point on agree to use the notation

$\|\underline{x} - \underline{a}\|$*

*Actually, if we use our notion of function as defined in Part 1 of this course, there is no need to talk about $\|\underline{x} - \underline{a}\|$. Rather, we could talk about $\|\underline{x}\|$, remembering that $\underline{x}$ merely represents the "input." In other words, from $\|\underline{x}\| = \sqrt{x_1{}^2 + \dots + x_n{}^2}$, we need only replace $\underline{x}$ by $\underline{x} - \underline{a}$ [$= (x_1 - a_1,\dots,x_n - a_n)$] to obtain $\|\underline{x} - \underline{a}\| = \sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2}$. However, to stress the fact that we are interested in the distance between $\underline{x}$ and $\underline{a}$, we shall usually use the form $\|\underline{x} - \underline{a}\|$.

rather than

$d(\underline{x},\underline{a})$.

In summary, we are simply defining $\|\underline{x} - \underline{a}\|$ to mean $\sqrt{(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2}$. (Notice that this definition of "distance" is really a function which maps vectors (n-tuples) into non-negative real numbers.)

All that remains for us to investigate in this section is whether our generalized "distance" concept has the usual algebraic properties that are associated with the geometric notion of distance. For the sake of simplicity, we shall limit our discussion to $\|\underline{x}\|$ rather than $\|\underline{x} - \underline{a}\|$,* since any results we obtain in one case are immediately applicable to the other.

Recall that in the case of "arrows," magnitude was defined in terms of absolute values, and that the properties of absolute values which we used in our computations were:

$$|\vec{x}| \geqslant 0 \text{ and } |\vec{x}| = 0 \text{ if and only if } \vec{x} = \vec{0} \tag{13}$$

$$|a\vec{x}| = |a|\ |\vec{x}|, \text{ where a is any real number} \tag{14}$$

$$|\vec{x} + \vec{y}| \leqslant |\vec{x}| + |\vec{y}| \tag{15}$$

What we would now like to show is that in E^n

$$\|\underline{x}\| \geqslant 0 \text{ and } \|\underline{x}\| = 0 \text{ if and only if } \underline{x} = \underline{0} \tag{13'}$$

$$\|a\underline{x}\| = |a|\ \|\underline{x}\|, \text{ where a is any scalar**} \tag{14'}$$

$$\|\underline{x} + \underline{y}\| \leqslant \|\underline{x}\| + \|\underline{y}\| \tag{15'}$$

In our demonstration, we shall again pick the special case $n = 4$, but it is hoped that you will see that the approach works for all values of n.

*The terminology parallels the lower dimensional cases. We call $\|\underline{x}\|$ the magnitude of $\underline{x}$ while the magnitude of $\underline{x} - \underline{a}$ (i.e., $\|\underline{x} - \underline{a}\|$) is called the distance between $\underline{x}$ and $\underline{a}$.

**Notice that since a is a scalar, we write $|a|$, not $\|a\|$. Furthermore, a cannot be a vector since we have not yet defined $\underline{a}\ \underline{x}$. In fact, we haven't even attempted as yet to generalize the dot and cross products to n-space.

(1) Is it true that $\|\underline{x}\| \geqslant 0$ and $\|\underline{x}\| = 0 \leftrightarrow \underline{x} = \underline{0}$?

Well,

$$\|\underline{x}\| = \sqrt{x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} + x^4}$$

and this is positive by definition of the (principle) square root of a non-negative number.

As for $\|\underline{x}\| = 0$, this implies that

$$x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} + x_4^{\ 2} = 0$$

but $x_1^{\ 2}$, $x_2^{\ 2}$, $x_3^{\ 2}$, and $x_4^{\ 2}$ are all non-negative, hence, the sum is 0 if and only if $x_1 = x_2 = x_3 = x_4 = 0$.

But, if $\underline{x} = (x_1,x_2,x_3,x_4)$ and if $x_1 = x_2 = x_3 = x_4 = 0$, then $\underline{x} = (0,0,0,0)$ which is $\underline{0}$.

(2) Is $\|a\underline{x}\| = |a|\ \|\underline{x}\|$ where a is a scalar?

Well,

$$\begin{aligned} |a\underline{x}| &= |(ax_1,ax_2,ax_3,ax_4)| \\ &= \sqrt{(ax_1)^2 + (ax_2)^2 + (ax_3)^2 + (ax_4)^2} \\ &= \sqrt{a^2(x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} + x_4^{\ 2})} \end{aligned}$$

Therefore,

$$\begin{aligned} \|a\underline{x}\| &= \sqrt{a^2}\ \sqrt{x_1^{\ 2} + x_2^{\ 2} + x_3^{\ 2} + x_4^{\ 2}} \\ &= |a|\ \|\underline{x}\|. \end{aligned}$$

(3) Finally, is it true that $\|\underline{x} + \underline{y}\| \leqslant \|\underline{x}\| + \|\underline{y}\|$?

This is a "toughie," computationally speaking. That is, let $\underline{x} = (x_1,x_2,x_3,x_4)$ and $\underline{y} = (y_1,y_2,y_3,y_4)$. Then we are being asked to show that

$$\sqrt{(x_1+y_1)^2 + (x_2+y_2)^2 + (x_3+y_3)^2 + (x_4+y_4)^2} \leqslant \sqrt{x_1^2+x_2^2+x_3^2+x_4^2}$$

$$+ \sqrt{y_1^2+y_2^2+y_3^2+y_4^2}.$$

Since both sides are non-negative, we may compare the squares of each side. Thus, we must check whether

$$(x_1+y_1)^2 + (x_2+y_2)^2 + (x_3+y_3)^2 + (x_4+y_4)^2 \leqslant (x_1^2+x_2^2+x_3^2+x_4^2)$$

$$+ 2\sqrt{x_1^2+x_2^2+x_3^2+x_4^2}$$

$$\sqrt{y_1^2+y_2^2+y_3^2+y_4^2}$$

$$+ (y_1^2+y_2^2+y_3^2+y_4^2)$$

Cancelling the appropriate terms reduces our inequality to

$$x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 \leqslant \sqrt{x_1^2+x_2^2+x_3^2+x_4^2}\ \sqrt{y_1^2+y_2^2+y_3^2+y_4^2}.$$

This result, which happens to be true (we leave the proof as a note for the end of this section), is known as <u>Schwarz's inequality</u>.

Schwarz's inequality, in one form or another, has application in many different topics. One topic that is of interest to us concerns the notion of a <u>dot product</u> in n-dimensional space. Based on our experience with lower dimensional spaces, it would seem natural to define

$$\underline{x}\cdot\underline{y}$$

as

$$x_1y_1 + \dots + x_ny_n \tag{16}$$

where

$$\underline{x} = (x_1,\dots,x_n) \text{ and } \underline{y} = (y_1,\dots,y_n).$$

The structural properties of dot products in the lower dimensional cases that made them useful to us were:

(i) $\underline{x}\cdot\underline{y} = \underline{y}\cdot\underline{x}$

(ii) $\underline{x}\cdot(\underline{y} + \underline{z}) = \underline{x}\cdot\underline{y} + \underline{x}\cdot\underline{z}$

(iii) $(c\underline{x})\cdot\underline{y} = c(\underline{x}\cdot\underline{y}) = \underline{x}\cdot(c\underline{y})$

(iv) $|\underline{x}\cdot\underline{y}| \leqslant \|\underline{x}\|\,\|\underline{y}\|$ (i.e., $\underline{x}\cdot\underline{y} = \|\underline{x}\|\,\|\underline{y}\| \cos\angle(\underline{x},\underline{y})$)

Now, it is almost trivial to verify that (i), (ii), and (iii) hold directly in E^n from the definition given in (16). Verifying (iv) is not as trivial, but it should be noticed that the validity of (iv) is precisely the statement of Schwarz's inequality, since $|\underline{x}\cdot\underline{y}|$ is $|x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4|$ while $\|\underline{x}\|\,\|\underline{y}\|$ is $\sqrt{x_1^2+x_2^2+x_3^2+x_4^2}\sqrt{y_1^2+y_2^2+y_3^2+y_4^2}$.

In fact, (iv) allows us to extend the concept of <u>angle</u> beyond 3-dimensional space. Namely, from (iv), it is clear that

$$\frac{|\underline{x}\cdot\underline{y}|}{\|\underline{x}\|\,\|\underline{y}\|} \leqslant 1.$$

(This is a trivial step in the sense that both sides of the inequality in (iv) are <u>numbers</u>, and for any positive numbers a and b, $a \leqslant b \leftrightarrow \frac{a}{b} \leqslant 1$.)

Thus, if we <u>define</u> the cosine of the angle between $\underline{x}$ and $\underline{y}$ (even if we can't picture it) to be $\frac{\underline{x}\cdot\underline{y}}{\|\underline{x}\|\,\|\underline{y}\|}$ then by this definition, the cosine has its usual property that $-1 \leqslant \cos\angle(\underline{x},\underline{y}) \leqslant 1$. We can then talk about <u>directional cosines</u> of vectors in n-space and we can even talk about two n-tuples being <u>orthogonal</u> (perpendicular) in terms of their dot product being zero. These ideas are pursued at great length in many real applications of mathematics (for example, the study of orthogonal functions [such as are used in Fourier Series] which we shall mention in a little more detail in Block 7), but for now we only hope that you begin to get the feeling that n-dimensional space is as real for n > 3 as it is for $n \leqslant 3$, and that, at least from an analytical point of view, distance is as real in the higher dimensional spaces as it is in the lower ones - even though we may have trouble at first trying to feel at home with the idea.

As a final remark, our definition of $\|\underline{x}\|$ leads to an extension of distance as we know it in ordinary geometry. For this reason, the definition

$$\|\underline{x}\| = \|(x_1,\ldots,x_n)\| = \sqrt{x_1^2 + \ldots + x_n^2}$$

is known as the Euclidean Metric (where Euclidean indicates ordinary geometry and metric indicates a measure of distance*).

A NOTE ON THE PROOF OF SCHWARZ'S INEQUALITY

There are probably many ways of proving Schwarz's inequality. One of the most elegant, if not the most obvious, makes use of elementary calculus.

Given the quadratic function defined by $f(x) = ax^2 + 2bx + c$ with $a > 0$, we compute f'. When $f'(x)$ is 0, we have a minimum. In fact, $f'(x) = 2ax + 2b$ while $f''(x) = 2a$ which is positive. Thus, the minimum occurs at $x = -b/a$. When $x = -b/a$, we have that $f(x) = a(-b/a)^2 + 2b(-b/a) + c = c - b^2/a$. In other words, the minimum value for $ax^2 + bx + c$ is $c - b^2/a$. Therefore, as soon as $ac - b^2$ is non-negative, then, $ax^2 + 2bx + c$ must be at least as great as zero since its minimum value is.

In summary then, if $\underline{a}$ is positive, then $ax^2 + 2bx + c$ is at least as great as zero for all x if and only if $ac - b^2 \geqslant 0$.

With this result in mind, we now look at

$$f(x) = (a_1x + b_1)^2 + \ldots + (a_nx + b_n)^2. \tag{1}$$

Since f as defined in (1) is the sum of squares, it is non-negative. That is, $f(x) \geqslant 0$ for all x. On the other hand, we can rewrite (1) as

$$f(x) = (a_1^2 + \ldots + a_n^2)x^2 + 2(a_1b_1 + \ldots + a_nb_n)x + (b_1^2 + \ldots + b_n^2). \tag{2}$$

*In terms of our earlier remarks, a function $d: E^n \to E$ is called a metric (distance function) on E^n if and only if

(i) $d(\underline{x}) \geqslant 0$ for all $\underline{x} \in E^n$ and $d(\underline{x}) = 0 \leftrightarrow \underline{x} = \underline{0}$

(ii) $d(c\underline{x}) = |c|d(\underline{x})$ for all real numbers c

(iii) $d(\underline{x} + \underline{y}) \leqslant d(\underline{x}) + d(\underline{y})$

Notice that, as written in (2), f has the form we just mentioned. Namely,

$$f(x) = ax^2 + 2bx + c$$

where

$$a = (a_1{}^2 + \ldots + a_n{}^2) \geqslant 0$$

$$b = (a_1 b_1 + \ldots + a_n b_n)$$

$$c = (b_1{}^2 + \ldots + b_n{}^2).$$

Moreover, as seen from (1), $f(x) \geqslant 0$ for all x. Consequently, our previous result guarantees that $b^2 \leqslant ac$. That is,

$$(a_1 b_1 + \ldots + a_n b_n)^2 \leqslant (a_1{}^2 + \ldots + a_n{}^2)(b_1{}^2 + \ldots + b_n{}^2).$$

By taking square roots, Schwarz's inequality follows.

Again, It is not our purpose here to make it seem that the proof of the inequality is either easy or natural. It is just for the sake of completeness that we felt obligated to present a proof. This might also serve as a good illustration as to why one must, upon occasion, solve abstract problems to understand better a real situation. In other words, it is possible that a problem involving either Schwarz's inequality or its proof might hardly seem an inspiring exercise in its own right if, for example, we had presented it as an exercise on max-min theory earlier in the course without the motivation afforded in the present context.

D

Limits

We are now in a position to talk more analytically about what we mean when we say that $\lim_{\underline{x} \to \underline{a}} f(\underline{x}) = L$.

Certainly, from an intuitive point of view, we would be tempted to say that $f(\underline{x})$ is near L if $\underline{x}$ is sufficiently close to $\underline{a}$. The point is that up until the previous section, the notion of $\underline{x}$ being close to $\underline{a}$ was developed only for 1, 2, and 3-dimensional space in terms of the arrow interpretation of vectors. From our discussion in the previous section, however, we can now extend the notion of "closeness" to n-dimensional space simply by defining "$\underline{x}$ is within δ of $\underline{a}$" to mean that $\|\underline{x} - \underline{a}\| < \delta$.

If we now mimic the formal definition of a limit as given for a real function of a single real variable, and recall only that absolute value for vectors (n-tuples) has been replaced by the notion of the Euclidean metric, we obtain the definition: We say that $\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = L$ if for every $\varepsilon > 0$ we can find $\delta > 0$ such that $0 < \|\underline{x} - \underline{a}\| < \delta$ implies that $|f(\underline{x}) - L| < \varepsilon$.

Perhaps the best way to illustrate our above discussion is by means of a specific example. In order to be able to capture the geometric significance of any remarks we make, we shall take the very special case for which the domain of f is E^2 (2-dimensional vector space). The point is that while all our remarks will then have geometric interpretation, they will also be self-contained from an arithmetical point of view. In this way, we can extend our results analytically from n = 2 to any general value of n and, at the same time, form a mental picture of what these results mean for lower dimensional cases in terms of usual geometry. In this sense, the most difficult job in studying real functions of n real variables is to handle the case n = 2 analytically, for once this is done our results will extend almost immediately to any number of real variables.

Let us then consider the example

$$f(\underline{x}) = f(x,y)^* = x^2 + y^3$$

where we wish to compute

$$\lim_{\underline{x}\to\underline{a}} f(\underline{x}),$$

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

If we rewrite this in traditional form, we have

*In deference to traditional mathematics, it is customary to denote 1-tuples by x rather than (x_1), 2-tuples by (x,y) rather than (x_1,x_2), and 3-tuples by (x,y,z) rather than (x_1,x_2,x_3).

$$\lim_{(x,y)\to(2,3)} (x^2 + y^3), \text{ or } \lim_{\substack{x\to 2\\ y\to 3}} (x^2 + y^3).* \tag{17}$$

Now, if we allow our intuition to reign, we sense that (17) "sort of" asks us to compute $x^2 + y^3$ when $x = 2$ and $y = 3$, from which we would guess that

$$\lim_{\substack{x\to 2\\ y\to 3}} (x^2 + y^3) = 31,$$

or in the language of the original exercise,

$$\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = 31. \tag{18}$$

The point is that (18) is at best a conjecture. We used our intuition in arriving at it, and we took certain liberties (such as letting $x = 2$ and $y = 3$) that led us to grief even in the simpler 1-dimensional case (i.e., Part 1 of our course). To test the validity of (18), we must use our "official" (rigorous) definition of limit to show that given any $\varepsilon > 0$ we can find $\delta > 0$ such that

$$0 < \|(x,y) - (2,3)\| < \delta$$

implies that

$$|x^2 + y^3 - 31| < \varepsilon,$$

or, by the definition of the Euclidean metric,

$$0 < \sqrt{(x - 2)^2 + (y - 3)^2} < \delta$$

implies that

$$|x^2 + y^2 - 31| < \varepsilon.$$

*We do not want to get involved with too many new concepts at one time, so we shall wait until the next section before going into more detail. However, the notation $(x,y)\to(2,3)$ or $\begin{cases} x\to 2\\ y\to 3\end{cases}$ presents many problems which do not at first meet the eye.

To take advantage of the fact that we want x to be near 2 and y to be near 3, we rewrite $x^2 + y^3 - 31$ in the more suggestive form

$(x^2 - 4) + (y^3 - 27)$.

Then,

$$|x^2 + y^3 - 31| = |(x^2 - 4) + (y^3 - 27)| \leqslant |x^2 - 4| + |y^3 - 27|. \tag{19}$$

From (19), we see that to have $|x^2 + y^3 - 31| < \varepsilon$, it is <u>sufficient</u> to be sure that $|x^2 - 4| + |y^3 - 27| < \varepsilon$.

But, $|x^2 - 4| + |y^3 - 27| < \varepsilon$ is guaranteed to happen if we can be sure that

$$\left.\begin{array}{l} |x^2 - 4| < \frac{\varepsilon}{2} \\ \text{and} \\ |y^3 - 27| < \frac{\varepsilon}{2}. \end{array}\right\} \tag{20}$$

The beauty of (20) is that each of the inequalities involves but a single real variable, and this is what we learned to handle in Part 1 of our course. For example, we are sure that we can find δ_1 such that

$$0 < |x - 2| < \delta_1 \rightarrow |x^2 - 4| < \frac{\varepsilon}{2} \tag{21}$$

since this was the definition of $\lim_{x\to 2} x^2 = 4$.

Similarly, from the fact that we know $\lim_{y\to 3} y^3 = 27$, we can find δ_2 such that

$$0 < |y - 3| < \delta_2{}^{*} \rightarrow |y^3 - 27| < \frac{\varepsilon}{2}. \tag{22}$$

*The type of computation for determining the δ_1 and δ_2 of (21) and (22) for a given $\frac{\varepsilon}{2}$ was done in detail in Part 1 and will not be repeated here so that we may continue uninterrupted. Since you may be a bit "rusty" with the computation, it is included in the study guide as Exercise 3.1.3(L). The interested reader may digress, if he so desires, and look at this exercise now.

Thus, for a given $\varepsilon > 0$, we can find δ_1 and δ_2 such that (21) and (22) both hold.

If we now let $\delta = \min\{\delta_1, \delta_2\}$ (i.e., $\delta \leqslant \delta_1$ and $\delta \leqslant \delta_2$), it follows from (21) and (22) that

$$0 < |x - 2| < \delta \text{ \underline{and} } 0 < |y - 3| < \delta \rightarrow |x^2 - 4| \text{ and } |y^3 - 27|$$

are both less than $\frac{\varepsilon}{2}$. That is,

$$\left.\begin{array}{l} 0 < |x - 2| < \delta \\ \\ 0 < |y - 3| < \delta \end{array}\right\} \rightarrow |x^2 - 4| + |y^3 - 27| < \varepsilon.$$

Hence, by (19),

$$\left.\begin{array}{l} 0 < |x - 2| < \delta \\ \\ 0 < |y - 3| < \delta \end{array}\right\} \rightarrow |x^2 + y^3 - 31| < \varepsilon. \tag{23}$$

The only trouble with (23) is that, while it seems to do the job for us, it does not have the right form. Namely, the δ we were seeking was to have the property that

$$0 < \sqrt{(x - 2)^2 + (y - 3)^2} < \delta \rightarrow |x^2 + y^3 - 31| < \varepsilon.$$

However, it is easy to see that

$$|x - 2| \leqslant \sqrt{(x - 2)^2 + (y - 3)^2} \quad \text{and} \quad |y - 3| \leqslant \sqrt{(x - 2)^2 + (y - 3)^2}\,*$$

Hence, if $\sqrt{(x - 2)^2 + (y - 3)^2} < \delta$ so also are $|x - 2|$ and $|y - 3|$. Thus, with δ as in (23), we have that

$$0 < \sqrt{(x - 2)^2 + (y - 3)^2} < \delta \rightarrow |x^2 + y^3 - 31| < \varepsilon \tag{24}$$

*That is, since $x_1{}^2 + \ldots x_n{}^2$ is a sum of non-negative numbers, the sum must be at least as great as each of the individual summands. That is, for $k = 1,\ldots,n$, $x_k{}^2 \leqslant x_1{}^2 + \ldots + x_n{}^2$. Hence, $|x_k| \leqslant \sqrt{x_1{}^2 + \ldots + x_n{}^2}$. In our present context, neither $(x - 2)^2$ nor $(y - 3)^2$ can exceed $(x - 2)^2 + (y - 3)^2$.

and we have, therefore, established the validity of

$$\lim_{(x,y)\to(2,3)} (x^2 + y^3) = 31.$$

Before continuing further, let us make the following aside. While our definition of limit required that we convert (23) into (24), it seems that in a major sense the job was done when we got to (23).

Now, in terms of our game of mathematics, let us recall that our theorems are no stronger than our assumptions. Thus, when we invented the Euclidean metric, it was only because we wanted to mimic our definition of distance in the lower dimensions. While this was a natural way to feel, there was no law that said we had to feel this way. For instance, suppose we wanted a metric only for finding limits. Then what (23) told us was that for $\underline{x}$ to be sufficiently close to $\underline{a}$, it was only necessary that x be sufficiently close to 2 and y sufficiently close to 3. More generally, if $\underline{x} = (x_1,\ldots,x_n)$ and $\underline{a} = (a_1,\ldots,a_n)$ then to make sure that $\underline{x}$ is within δ of $\underline{a}$ it is only necessary to make sure that the <u>maximum</u> of the numbers

$$|x_1 - a_1|,\ldots, \text{ and } |x_n - a_n|$$

is "sufficiently small."

For this reason, one often uses a different metric than the Euclidean one when one deals with limits. (Recall that "metric" means a "distance function.") In particular, in this case, one defines a metric called the <u>Minkowski metric</u> by

$$\|\underline{x}\|^* = \max\{|x_1|,\ldots,|x_n|\}. \qquad (25)$$

The important point is that the Minkowski metric also satisfies properties (13'), (14'), and (15') (otherwise it wouldn't be called a metric) and behaves as distance should behave. The proof is left for the exercises.

*Technically speaking, we should not use the same symbol for two different metrics. Rather than invent still new symbolism at this time, we shall use the same symbol for both the Euclidean and the Minkowski metric, but indicate each time which metric we are using. In a little while, we shall see that as far as limit problems are concerned, there is no harm in using the two metrics interchangeably.

Let us return now to the main stream of the present problem and recall that we chose $\underline{x}$ in E^2 so that we could exploit the geometric aspects of the problem.

When we look at $f(x,y) = x^2 + y^3$ graphically, we are talking about the surface $z = x^2 + y^3$. (There is really no need to have to know what this surface looks like in order to understand the following discussion. If, however, you feel ill at ease with the graphs of surfaces, the material contained, for example, in Thomas Sections 12.10 and 12.11 might be of interest to you.) At the point (2,3), the height of the surface above that point is 31. Pictorially,

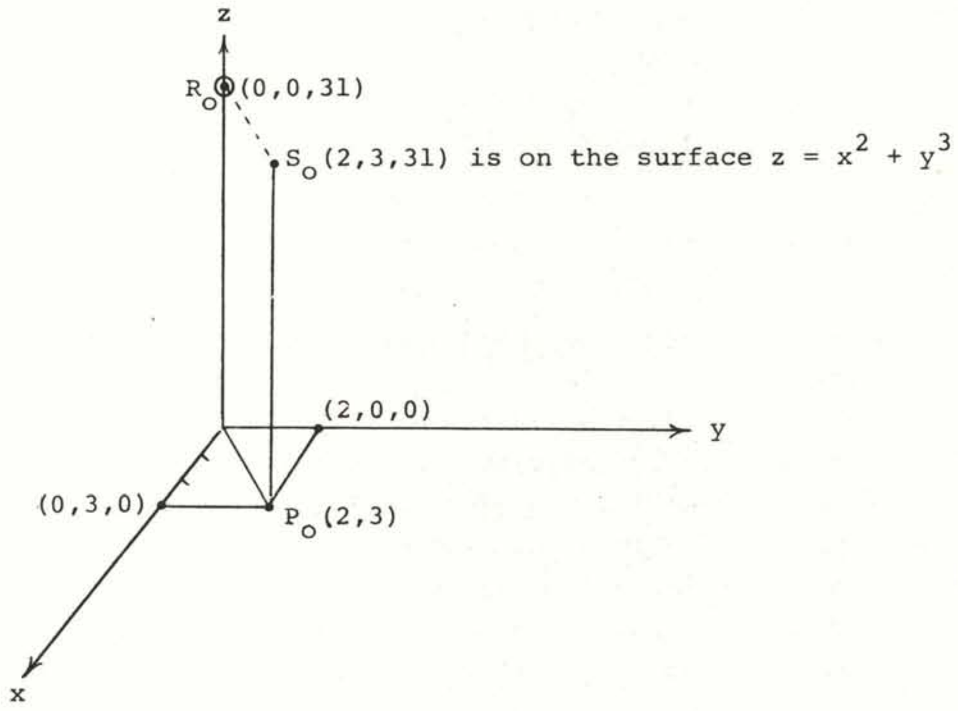


Figure 1

What our limit problem then asks is: How close must the point (x_1,y_1) in the xy-plane be to the point $P_o(2,3)$ if $f(x,y)$ is to be within ε of 31, i.e., if the height of the surface $z = x^2 + y^2$ above (x_1,y_1) is to be within ε of the height of the surface above (2,3)? Again, pictorially,

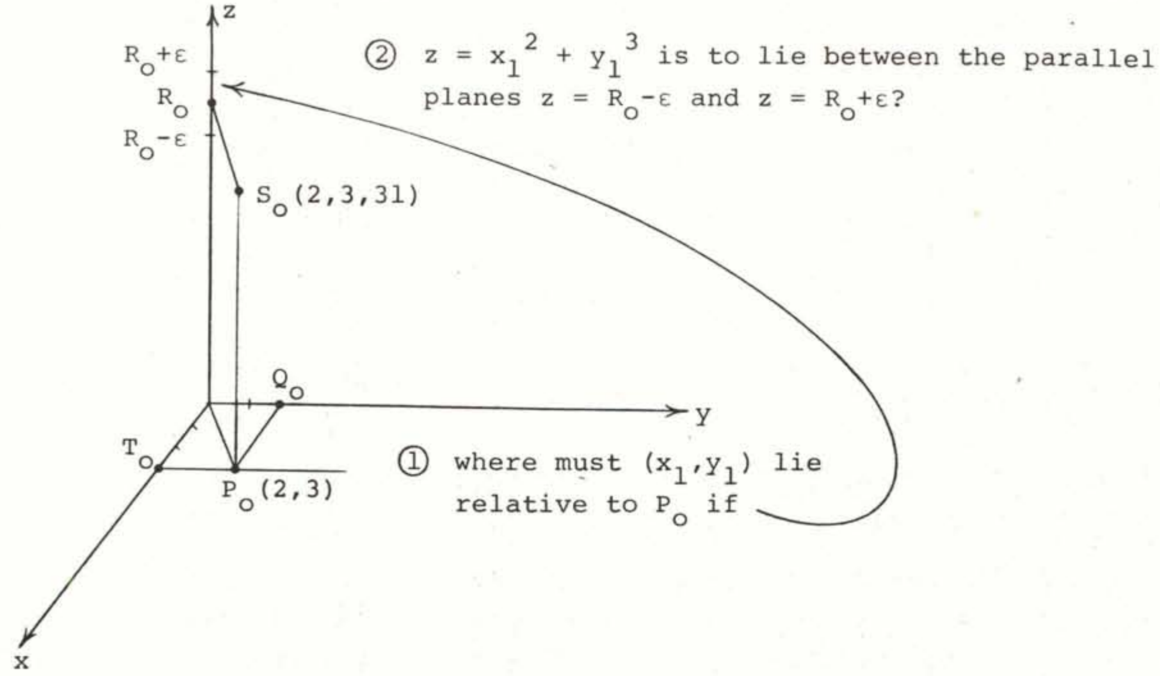
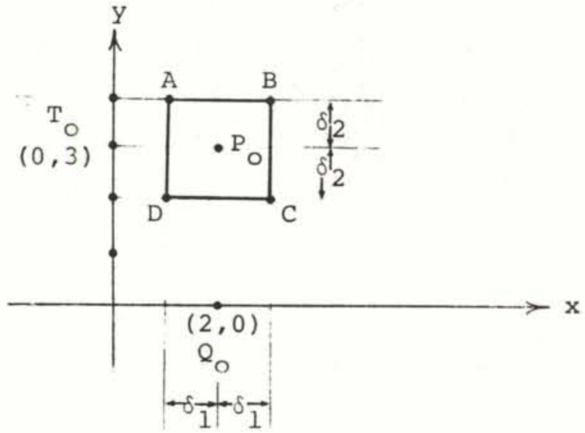


Figure 2

In terms of δ_1 and δ_2 as defined in (21) and (22), we have



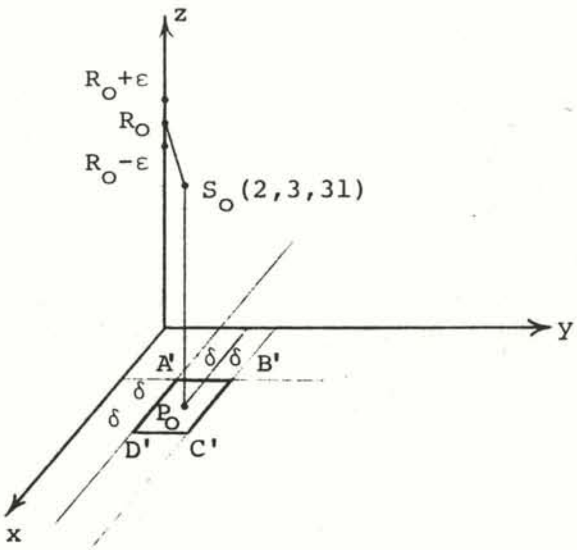
① The interior of rectangle ABCD consists of those points (x,y) in the xy-plane for which $|x - 2| < \delta_1$ and $|y - 3| < \delta_2$ (i.e., x is within δ_1 of 2 and y is within δ_2 of 3).

② For every point (x,y) inside ABCD, $31-\varepsilon < x^2 + y^3 < 31+\varepsilon$ (where we omitted the absolute value signs since near (2,3) $x^2 + y^3$ is obviously positive).

Figure 3

What the Minkowski metric does when we let $\delta = \min\{\delta_1,\delta_2\}$ is that it replaces ABCD of Figure 3 by a square whose side has length 2δ and whose center is P_o.

Again, pictorially,



For each (x,y) in the square A'B'C'D', the point $(x,y,x^2 + y^3)$ lies between the planes $z = 31-\varepsilon$ and $z = 31+\varepsilon$.

Figure 4

This discussion serves as an excellent springboard for our comparing the Euclidean and the Minkowski metrics. In particular, we can now see what it means when we say that the two metrics are equivalent with respect to limit problems.

Perhaps the easiest way to motivate our meaning of "equivalent" is to recall the definition of

$$\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = L.$$

We said that this meant for any $\varepsilon > 0$ we could find $\delta > 0$ such that

$$0 < \|\underline{x} - \underline{a}\| < \delta$$

implied that

$$|f(\underline{x}) - L| < \varepsilon.$$

The problem is that we do not know whether $\|\underline{x} - \underline{a}\|$ refers to the Minkowski metric or the Euclidean metric. Our claim is that it makes no difference! That is, whether or not it is true that $\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = L$ is independent of which metric we use in our definition of a limit.

This is not too difficult to see from an analytic point of view for any dimensional space, but in the case where we can draw the picture (as in the present example) we have a particularly simple geometrical interpretation.

For example, suppose for a certain $\varepsilon > 0$ we have exhibited a $\delta > 0$ such that for $0 < \|\underline{x} - \underline{a}\| < \delta$, $|f(\underline{x}) - L| < \varepsilon$. We then are told that we are using the Euclidean metric. To remind us of this, let us replace δ by δ_E.

The question then becomes: Does there exist another (or perhaps the same) δ (say δ_M) such that $0 < \|\underline{x} - \underline{a}\| < \delta_M \rightarrow |f(\underline{x}) - L| < \varepsilon$ where now $\|\underline{x} - \underline{a}\|$ is the Minkowski metric?

In other words, given that

$$0 < \sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2} < \delta_E \rightarrow |f(\underline{x}) - L| < \varepsilon,$$

find $\delta_M > 0$ such that

$$0 < \max\{|x_1 - a_1|, \dots, |x_n - a_n|\} < \delta_M \rightarrow |f(\underline{x}) - L| < \varepsilon.$$

Our claim is that we need only choose

$$\delta_M = \frac{\delta_E}{\sqrt{n}} .$$

For if $\max\{|x_1 - a_1|, \dots, |x_n - a_n|\}$ is less than $\frac{\delta_E}{\sqrt{n}}$, then each of the quantities, $|x_1 - a_1|, \dots, |x_n - a_n|$ is less than $\frac{\delta_E}{\sqrt{n}}$. Therefore, $(x_1 - a_1)^2, \dots,$ and $(x_n - a_n)^2$ are each less than $\frac{\delta_E^{\,2}}{n}$, whereupon

$$\sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2} < \sqrt{\underbrace{\frac{\delta_E^{\,2}}{n} + \dots + \frac{\delta_E^{\,2}}{n}}_{\text{n times}}} = \sqrt{\frac{n\ \delta_E^{\,2}}{n}} = \delta_E.$$

But, since $\sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2} < \delta_E$, we know from above that $|f(\underline{x}) - L| < \varepsilon$.

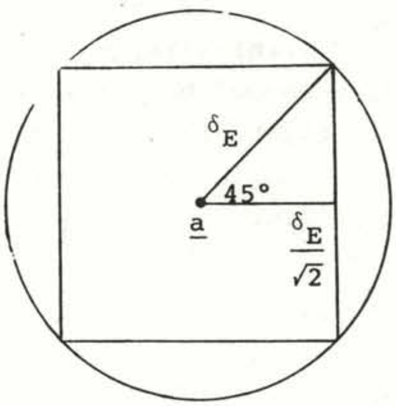
In the case n = 2, this says that $\delta_M = \frac{\delta_E}{\sqrt{2}}$. What does this mean pictorially? Well, saying that

$$\sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2} < \delta_E \rightarrow |f(\underline{x}) - L| < \varepsilon$$

means

① If $\underline{x}$ is in the circle, $|f(\underline{x}) - L| < \varepsilon$.

② Hence, if $\underline{x}$ is in the square which is inscribed in the circle, $|f(\underline{x}) - L| < \varepsilon$ (since the square is contained in the circle).



③ The "half-side" of the inscribed square is $\frac{\delta_E}{\sqrt{2}}$.

④ Thus, in terms of the Minkowski metric, for every $\underline{x}$ in the square $\|\underline{x} - \underline{a}\| < \frac{\delta_E}{\sqrt{2}}$ since $\max\{|x_1 - a_1|, |x_2 - a_2|\} < \frac{\delta_E}{\sqrt{2}}$.

In "plain English," this says that once we find a "circular" neighborhood of $\underline{a}$ with the desired property, we can always find a "square"* neighborhood of $\underline{a}$ with the same desired property.

Conversely, if we have found δ_M such that $0 < \max\{|x_1 - a_1|,\ldots, |x_n - a_n|\} < \delta_M \rightarrow |f(\underline{x}) - L| < \varepsilon$, our claim is that this is enough to insure that if

$$0 < \sqrt{(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2} < \delta_M$$

then

$$|f(\underline{x}) - L| < \varepsilon$$

(i.e., we may choose $\delta_E = \delta_M$).

*While the picture is easy to see when n = 2, the mathematician extends the concept of circles and squares to all dimensions. That is, we define an <u>n-sphere of radius δ</u> in E^n by $x_1{}^2 + \ldots + x_n{}^2 \leqslant \delta^2$, while we define an <u>n-cube of side 2a</u>, $|x_1| \leqslant a,\ldots,|x_n| < a$. What we have proven in general (and illustrated geometrically for n = 2), is that we can always "inscribe" an n-cube in an n-sphere.

Namely,

$$0 < \sqrt{(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2} < \delta_M \rightarrow$$

$$0 < (x_1 - a_1)^2 + \ldots + (x_n - a_n)^2 < \delta_M{}^2.$$

Then since $(x_k - a_k)^2$ $(k = 1,\ldots,n) \leqslant (x_1 - a_1)^2 + \ldots + (x_n - a_n)^2$ [because the left side is but one of the terms which makes up the sum of non-negative terms on the right side], it follows that

$$(x_k - a_k)^2 < \delta_M{}^2, \text{ for } k = 1,\ldots,n.$$

Therefore,

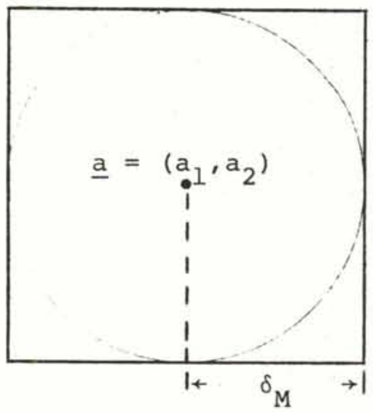
$$|x_k - a_k| < \delta_M{}^*, \; k = 1,\ldots,n.$$

Since $|x_k - a_k| < \delta_M$ for each $k = 1,\ldots,n$, it follows that

$$\max\{|x_1 - a_1|,\ldots,|x_n - a_n|\} < \delta_M$$

and this insures that $|f(\underline{x}) - L| < \varepsilon$.

In terms of a picture for $n = 2$



$\max\{|x_1 - a_1|,|x_2 - a_2|\} < \delta_M$ requires only that we be inside the square.

Thus, being inside the inscribed circle (whose radius is δ_M) guarantees that we are within the square.

*Recall that $a^2 < b^2 \rightarrow |a| < |b|$, not $a < b$. For example, $(-3)^2 < (-5)^2$ even though $-3 > -5$.

Notice that while in this case $\delta_E = \delta_M$, this was not what was important. What is important is that given δ_M such that

$$0 < \max\{|x_1 - a_1|,\ldots,|x_n - a_n|\} < \delta_M \rightarrow |f(\underline{x}) - L| < \varepsilon$$

we can find δ_E such that

$$0 < \sqrt{(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2} < \delta_E \rightarrow |f(\underline{x}) - L| < \varepsilon$$

and conversely. In intuitive terms, we can inscribe n-spheres in n-cubes and n-cubes in n-spheres. As to when $\delta_E = \delta_M$ and when $\frac{\delta_E}{\sqrt{n}} = \delta_M$, perhaps the easiest way to keep the distinction in mind is pictorially in the case n = 2. That is, given a circular neighborhood, the circumscribed square contains points outside the sphere and hence the result need not be true for such points. Given the "square" neighborhood, every point within the inscribed circle also lies in the square.

Lest we have lost track of our original aim, what we have shown is that we can exhibit a Minkowski neighborhood of $\underline{a}$ such that for all $\underline{x}$ in this neighborhood $|f(\underline{x}) - L| < \varepsilon$ <u>if and only if</u> we can exhibit a Euclidean neighborhood of $\underline{a}$ such that for all $\underline{x}$ in this neighborhood $|f(\underline{x}) - L| < \varepsilon$.

It is not important that the two neighborhoods have the same size (whatever this is to mean). All that we need for determining a limit is to find for any given $\varepsilon > 0$ one δ-neighborhood with the desired property.

E

An Introduction to Continuity

In the last section, when we wrote, for example, $\lim\limits_{(x,y)\rightarrow(2,3)} (x^2 + y^3)$, there was a bit of subtlety that we preferred to ignore until now in the hope that we have had additional time to get used to metrics in n-space.

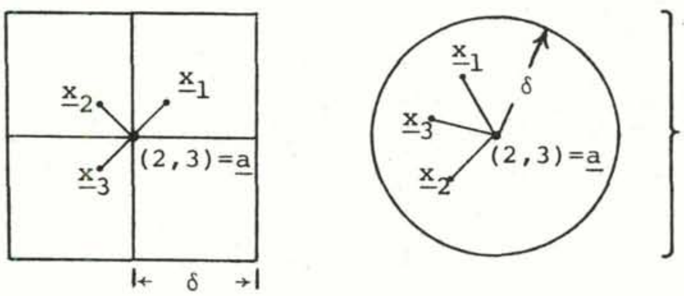
The point is just as in the calculus of a single real variable, when we say that $\lim\limits_{x\rightarrow a} f(x) = L$ we mean that the limit exists, and <u>in each case</u> is equal to L, <u>regardless of the "path" by which x approaches a</u>. The problem was not too severe in the 1-dimensional case, since there were only two ways in which x could approach a -- either through values less than a or through values greater than a. (Pictorially, the domain of x was the x-axis, so that x was either to the right of a or to the left of a.) In any case, we took this into account by saying that $\lim_{x \to a} f(x) = L$ implied that both $\lim_{x \to a^+} f(x)$ and $\lim_{x \to a^-} f(x)$ existed and that each was equal to L.

From an analytical point of view, the key was in the fact that when we wrote $0 < |x - a| < \delta$, all that we implied was that $a - \delta < x < a + \delta$, $(x \neq a)$, and as a result, we were taking into account that x could either exceed a or be less than a.

The same problem exists in n-dimensional vector spaces in general. For example, our definition that $\lim_{(x,y) \to (2,3)} (x^2 + y^3) = 31$ took into account the fact that not only did the limit exist, but its value in no way depended upon how we let (x,y) approach (2,3).

Pictorially, the problem is that the domain of our function f in this case is the xy-plane, and as a result, there are infinitely many paths (even if we restrict our attention to straight line approaches) by which (x,y) can approach (2,3). The way that we showed in our definition that the limit did not depend on the path was when we exhibited $\delta > 0$ such that $0 < \|(x,y) - (2,3)\| < \delta \to |x^2 + y^3 - 31| < \varepsilon$.

For, whether we used the Minkowski metric or the Euclidean metric, the implication was that $|x^2 + y^3 - 31| < \varepsilon$ for <u>every</u> value of x and y within a sufficiently small neighborhood of (2,3). Again, pictorially,



Once $\|\underline{x} - \underline{a}\| < \delta$, then $|x^2 + y^3 - 31| < \varepsilon$, regardless of the direction of $\underline{x}$ relative to $\underline{a}$.

In a similar way, even when we deal with n-dimensional space in general, the fact that we say $\|\underline{x} - \underline{a}\| < \delta$ implies that we are talking about <u>every</u> $\underline{x}$ within δ of $\underline{a}$.

We will leave computational illustrations to the exercises, but for now we want to make it clear that the definition of $\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = L$ must not depend on how we let $\underline{x}$ approach $\underline{a}$. (We shall refine these ideas in later chapters.)

Because this point is so important, we would like to restate it in still different language. In the case of 2 (or 3) dimensional space, what we mean when we say that $\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = L$ is that once given $\varepsilon > 0$, we can find a number $\delta > 0$ such that once $\underline{x}$ is in the circle (sphere) centered at $\underline{a}$ with radius equal to δ, or in the square (cube) of side 2δ centered at $\underline{a}$, then $f(\underline{x})$ is within ε of L. Since this is true for <u>every</u> point in the neighborhood of $\underline{a}$, the fact that $f(\underline{x})$ approaches L as $\underline{x}$ approaches $\underline{a}$ does not depend on the particular path we choose for allowing $\underline{x}$ to approach $\underline{a}$.

With this idea in mind, we now mimic our definition in the case of functions of a single real variable, and we define continuity as follows.

Let $f:E^n \to E$ and let $\underline{a} \in E^n$. Then we say that f is <u>continuous</u> at $\underline{a}$ if and only if $\lim_{\underline{x}\to\underline{a}} f(\underline{x}) = f(\underline{a})$. What this says is that for f to be continuous at $\underline{a}$, it must be that (1) f is defined at $\underline{a}$, i.e., $f(\underline{a})$ exists, and (2) no matter how $\underline{x}$ approaches $\underline{a}$, we can make $f(\underline{x})$ come as close in value as we wish to $f(\underline{a})$ just by choosing $\underline{x}$ close enough to $\underline{a}$.

For example, without saying it in as many words, we showed that if f was defined by $f(x,y) = x^2 + y^3$ then f was continuous at (2,3). That is, we showed that (1) f(2,3) existed and was equal to 31 and (2) $\lim_{(x,y)\to(2,3)} f(x,y) = f(2,3) = 31$ regardless of how (x,y) approached (2,3).

Pictorially, this meant that the surface $z = x^2 + y^3$ was unbroken in a neighborhood of the point (2,3,31).

We do not want to introduce too many new ideas at one time in our introductory discussion of functions of several variables, but we felt that you should be cautioned that many concepts that looked harmless in the 1-dimensional case become rather nasty when we consider more general dimensional spaces. In particular, the problem of letting $\underline{x}$ approach $\underline{a}$ in n-dimensional space offers a few pitfalls that were not so apparent when we dealt with the 1-dimensional case. Further remarks are left to the exercises.

F

A Note on n-dimensional Vector Spaces

In terms of arrows versus n-tuples, we probably have the feeling that if $n = 1$, 2, or 3, then we are dealing with arrows while if n exceeds 3 we are dealing with n-tuples. It is very important to notice at this time that even in the lower dimensional cases there are situations in which we feel more comfortable talking, say, about 2-tuples than about arrows in the plane.

For example, recall that while the rules may have been motivated by the arrow interpretation, the fact remains that a 2-dimensional vector space is defined by the set of all 2-tuples together with the following structure:

(i) $(a_1,a_2) = (b_1,b_2)$ means that $a_1 = b_1$ and $a_2 = b_2$

(ii) $(a_1,a_2) + (b_1,b_2) = (a_1 + b_1, a_2 + b_2)$

(iii) $r(a_1,a_2) = (ra_1,ra_2)$, for any scalar r.

Trivially, the set of planar vectors satisfy (i), (ii), and (iii). However, there are other, quite different, sets which also satisfy (i), (ii), and (iii). To this end, we shall look at one such set.

Consider the set of all polynomials of degree less than or equal to 1. That is, let $S = \{a_o + a_1x : a_o, a_1 \varepsilon R\}$ where R denotes the real numbers. We already know the following facts:

(1) $a_o + a_1x = b_o + b_1x \leftrightarrow a_o = b_o$ and $a_1 = b_1$

(2) $(a_o + a_1x) + (b_o + b_1x) = (a_o + b_o) + (a_1 + b_1)x$

(3) $r(a_o + a_1x) = (ra_o) + (ra_1)x$ for any $r \varepsilon R$.

Since the sum and product of real numbers are also real numbers, notice that the operations defined by (2) and (3) yield elements of S. That is, the sum of any two members of S and any constant multiple of a member of S are again members of S.*

Suppose we now decide to abbreviate $a_o + a_1x$ by the 2-tuple (a_o, a_1).** With this notation in mind, we see that properties (i), (ii), and (iii) are obeyed. That is, if we define equality, addition, and scalar multiplication for linear polynomials as we have, then S serves as a <u>real model</u> which also satisfies (i), (ii), and (iii). In still other words, if we look at vectors in the plane in terms of what we mean by vector equality, vector addition, and scalar multiplication, and we also look at all polynomials of degree no greater than 1 with respect to what we mean by polynomial equality, polynomial addition, and polynomial "scalar" multiplication (i.e., multiplication by a constant), then we cannot tell the difference between these two <u>different</u> structures. If this sounds like a contradiction, again what we're trying to say is that any differences between polynomials of degree no greater than 1 and vectors in the plane must be caused by properties which are independent of properties (i), (ii), and (iii). That is, in terms of our game of mathematics concept, since both models obey (i), (ii), and (iii), they will both obey any "facts" which follow inescapably from (i), (ii), and (iii). For example, if we take the derivative (in the usual sense) of a polynomial of degree less than or equal to 1, we wind up with the set of all real numbers as possible solutions. On the other hand, since our vectors in the plane are constants, their derivative will always be 0. The point is that the

---

*This is why we must let S include constants as well as first degree polynomials. For example, if S is restricted to consist of linear polynomials, i.e., all polynomials of the form $a_o + a_1x$ <u>where</u> $a_1 \neq 0$, then neither the sum nor a scalar multiple of members of S need be a member of S. As an example, notice that both $2x + 3$ and $-2x + 4$ are linear polynomials but the sum of the two is simply 7, which is a constant, <u>not</u> a linear polynomial. In a similar vein, 0 is a real number and $0(2x + 3) = 0$, which again is a constant but not a first degree (linear) polynomial. The key point is that by letting S consist of constants and first degree polynomials we wind up with the required "closure."

**Notice that the concept of a two-tuple does not depend on how the members of the pair are labeled. Thus, (a,b), (x,y), (a_1,a_2), (a_o,a_1),... are all acceptable notations. Since it is more natural to write a polynomial as $a_o + a_1x$ than as $a_1 + a_2x$, we used (a_o,a_1) as our 2-tuple rather than (a_1,a_2). The important thing is that the first member of the 2-tuple is being used to denote the constant term while the second member is being used to denote the coefficient of x. If you feel more comfortable using the notation (a_1,a_2) so that it agrees with our previous discussion, simply think of the polynomial as being $a_1 + a_2x$.

concept of derivative is independent of (i), (ii), and (iii) and hence we cannot "notice" this difference if we are restricted to studying (i), (ii), (iii), <u>and</u> their consequences.

In summary, what we are saying is that if we specify that we want S to be the set of all real valued 2-tuples subject to the three "rules":

(1) $(a,b) = (c,d) \leftrightarrow a = c$ and $b = d$

(2) $(a,b) + (c,d) = (a + c, b + d)$

(3) $r(a,b) = (ra,rb)$ for any $r \varepsilon R$,

then the set of planar vectors and the set of polynomials of degree less than or equal to 1 both serve as models for S with respect to the obvious operations in each case. [Moreover, any differences between the two models must stem from properties that are not inescapable consequences of (1), (2), and (3).] Yet, somehow or other we probably do not feel comfortable describing linear polynomials by arrows.

In terms of our structure of mathematics, this is what led us to a more general definition of a 2-dimensional vector space. We start with a <u>set</u> S of 2-tuples. That is, $S = \{(a,b): a$ and b are real numbers$\}$. So far, S is only a set, since we have imposed no structure on it. We now impose a structure on S as follows.

(1) We define an equivalence relation on S so that if $\underline{a}$* $= (a_1,a_2)$ and $\underline{b} = (b_1,b_2)$ are both members of S, we will agree that $\underline{a} = \underline{b}$** means that $a_1 = b_1$ <u>and</u> $a_2 = b_2$.

(2) We will define a binary operation on S such that with $\underline{a}$ and $\underline{b}$ as above, $\underline{a} + \underline{b}$*** shall mean that member of S named by $(a_1 + b_1, a_2 + b_2)$.

*We switch from $\vec{a}$ to $\underline{a}$ to make it clear that even a 2-tuple need not be viewed as an "arrow" in the plane. For example, $\underline{a}$ could denote $a_o + a_1x$.

**It is pedagogically dangerous to let the same symbol have two different meanings. Here we are using "=" to denote <u>two</u> equivalence relations, one between numbers and one between 2-tuples. Perhaps we should have used a different symbol, say $\circleddash$ to denote the equivalence of 2-tuples. For example, $\underline{a} \circleddash \underline{b}$ means $a_1 = a_2$ and $b_1 = b_2$ (where "=" is in the usual context since a_1, a_2, b_1, and b_2 are numbers). We hope that it will be clear from text whether = refers to numbers or 2-tuples.

***This is the same problem as above. Perhaps we should have written $\underline{a} \oplus \underline{b} = (a_1 + b_1, a_2 + b_2)$.

(3) If r is any real number and $\underline{a} = (a_1,a_2)$, the scalar product $r\underline{a}$ is defined by (ra_1,ra_2).

If the set S obeys rules (1), (2), and (3) [which, by the way, were motivated by properties (i), (ii), and (iii) of our previous two models] then we refer to this structure as being a 2-dimensional vector space and denote it by E^2. Notice that E^2 is the set S together with a structure which is defined by (1), (2), and (3).

We can now mimic (i), (ii), and (iii) [or (1), (2), and (3)] as we have already done in Section C for n-tuples in general, and in this way, we can invent the n-dimensional vector space.

By way of review, we let S_n denote the set of all n-tuples $(r_1,\ldots,r_n)$ where $r_1,\ldots,r_n$ are real numbers. Using the abbreviation $\underline{a} = (a_1,\ldots,a_n)$ to denote elements of S_n, we imposed the following structure on S_n:

(a) For $\underline{a}$ and $\underline{b}$ in S_n, we say that $\underline{a} = \underline{b}$ if and only if $a_1 = b_1,\ldots,$ and $a_n = b_n$.

(b) For $\underline{a}$ and $\underline{b}$ in S_n, we define $\underline{a} + \underline{b}$ by $(a_1 + b_1,\ldots,a_n + b_n)$.

(c) For $\underline{a}$ in S_n, if r is any real number, we define $r\underline{a}$ to mean $(ra_1,\ldots,ra_n)$.

Any set S_n together with the structure implied by (a), (b), and (c) is called an n-dimensional vector space, and is denoted usually by E^n.

What is most important to understand here is that while, geometrically speaking, it is harder to think of n-dimensional space when n exceeds three than it is to think of a 1, 2, or 3-dimensional vector space, analytically speaking, the meaningfulness of an n-dimensional space does not depend on the value of n.

As a specific example, consider the set S_4 of all polynomials of degree less than or equal to three. That is, an element of S_4 has the form

$$a_0 + a_1x + a_2x^2 + a_3x^3 \tag{7}$$

where a_0, a_1, a_2, and a_3 are all real numbers.

If we define equality, addition, and "scalar multiplication" of polynomials in the usual way, and if we agree to abbreviate (7) by $\underline{a}$, it is quickly verified that:

(a) If $\underline{a}$ and $\underline{b}$ belong to S_4, then $\underline{a} = \underline{b}$ if and only if $a_o = b_o$, $a_1 = b_1$, $a_2 = b_2$, and $a_3 = b_3$.

(b) If $\underline{a}$ and $\underline{b}$ belong to S_n, then $\underline{a} + \underline{b} = (a_o + b_o, \ldots, a_3 + b_3)$.

(c) If $\underline{a}$ belongs to S_4 and r is any real number, then $r\underline{a} = (ra_o, ra_1, ra_2, ra_3)$.

Thus, with respect to the required structure, the system of polynomials of degree less than or equal to three is a real model of a 4-dimensional vector space, as real a model as the structure of linear polynomials is of a 2-dimensional vector space.