MATRIX ALGEBRA IN THE STUDY OF FUNCTIONS OF SEVERAL VARIABLES

A

Introduction

In Block 3, we emphasized the form $f(\underline{x})$. Notice that if we elect to view the real number system as a 1-dimensional vector space (i.e., as the set of 1-tuples), then $f(\underline{x})$ may be viewed as a special case of $\underline{f}(\underline{x})$. In other words, we may identify

$$f:E^n \to R$$

with

$$\underline{f}:E^n \to E^1 .$$

Quite apart from this rather trivial identification, there are many important reasons for introducing the study of mapping vector spaces into vector spaces. The reason, we shall exploit in the development of Block 4, is that the concept of systems of equations in several unknowns (variables) lends itself very nicely to the format of studying functions from one vector space into another.

For example, rather than a single equation of the form $y = g(x_1,\ldots,x_n)$, suppose we had a system of m such equations, say

$$\left.\begin{array}{l} y_1 = g_1(x_1,\ldots,x_n) \\ \cdot \\ \cdot \\ \cdot \\ y_m = g_m(x_1,\ldots,x_n) \end{array}\right\} \qquad (1)$$

The point that we wish to exploit in this chapter is that the system of equations defined by (1) can be viewed as a <u>single</u> <u>vector</u> function which maps E^n into E^m. To be more specific, we may view equations (1) as the mapping which sends $(x_1,\ldots,x_n)$ into $(y_1,\ldots,y_m)$, where $y_1,\ldots,y_m$ are as given by equations (1). That is, we may define $\underline{g}:E^n \to E^m$ by

$$\underline{g}(x_1,\ldots,x_n) = [g_1(x_1,\ldots,x_n),\ldots,g_m(x_1,\ldots,x_n)]$$

In terms of a more concrete example, suppose we have that

$$\left.\begin{array}{l} y_1 = x_1^2 + 3x_2 + 4x_3 + 2x_4^3 \\ y_2 = x_1 - 4x_2^2 + 5x_3^2 + 6x_4 \end{array}\right\} \quad (2)$$

then we are saying that equations (2) induce a function $\underset{\sim}{g}: E^4 \to E^2$, where $\underset{\sim}{g}$ is defined by

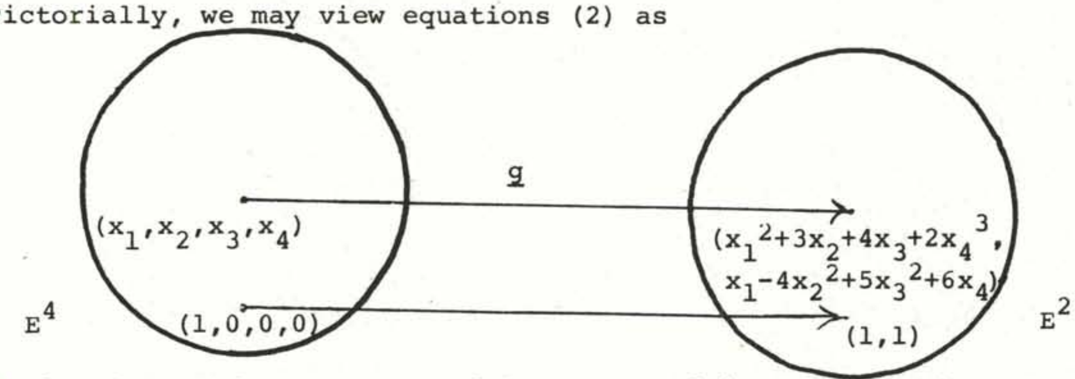
$$\underset{\sim}{g}(x_1,x_2,x_3,x_4) = (x_1^2 + 3x_2 + 4x_3 + 2x_4^3, x_1 - 4x_2^2 + 5x_3^2 + 6x_4). \quad (3)$$

For example, equation (3) yields

$$\underset{\sim}{g}(1,0,0,0) = (1,1) \quad (4)$$

where (4) was obtained from (3) by letting $x_1 = 1$ and $x_2 = x_3 = x_4 = 0$.

Pictorially, we may view equations (2) as



Notice that such concepts as 1-1, onto, and inverse functions were defined for mappings of any set into another. In particular, then, we may ask whether $\underset{\sim}{g}$ is 1-1 and/or onto, but such inquiries lead to extensive computational manipulations. For example, with respect to equation (3), asking whether $\underset{\sim}{g}$ is onto is equivalent to the algebraic problem:

> Given any pair of real numbers y_1 and y_2, do there exist values for x_1, x_2, x_3, and x_4 such that
>
> $$y_1 = x_1^2 + 3x_2 + 4x_3 + 2x_4^3$$
>
> $$y_2 = x_1 - 4x_2^2 + 5x_3^2 + 6x_4\ ?$$

Notice that this is again (2), with a different emphasis.

In this context, equation (4) says that one solution of (2) if $y_1 = y_2 = 1$ is $x_1 = 1$ and $x_2 = x_3 = x_4 = 0$. There may, of course, be other solutions, in which case $\underline{g}$ would not be 1-1.

Notice also that (2) is a fairly simple illustration of equations (1). In general, equations (1) may be so "messy" that the algebra becomes very difficult if not completely hopeless.

There is, of course, one particularly simple choice of $g_1,\ldots,g_m$ in equations (1) that is easy to handle algebraically, and that is the case in which the g's are linear combinations of $x_1,\ldots,x_n$. By a linear combination of $x_1,\ldots,$ and x_n, we mean any expression of the form

$$a_1x_1 + \ldots + a_nx_n$$

where $a_1,\ldots,$ and a_n are constants.

In fact, the first equations we learned to handle in our elementary algebra courses were of this type. All too soon we were taught that these were overly-simple, whereupon we moved on to "harder" equations such as quadratics, cubics, logarithms, etc. Yet the astounding point is that, in the midst of some very serious complex arithmetic which surrounds the calculus of several variables, we find that the "simple" linear combinations are the backbone of our investigations. More specifically, we have seen that if f is a differentiable function of $\underline{x} = (x_1,\ldots,x_n)$ at $\underline{x} = \underline{a}$ (that is, if $f_{x_1},\ldots,f_{x_n}$ all exist and are continuous at $\underline{x} = \underline{a}$) then

$$\Delta f = f(a_1 + \Delta x_1,\ldots,a_n + \Delta x_n) - f(a_1,\ldots,a_n) = \{f_{x_1}(a_1,\ldots,a_n)\Delta x_1 + \ldots + f_{x_n}(a_1,\ldots,a_n)\Delta x_n\} + [k_1\Delta x_1 + \ldots + k_n\Delta x_n] \tag{5}$$

where $k_1,\ldots,k_n \to 0$ as $\Delta x_1,\ldots,\Delta x_n \to 0$.

Since the bracketed expression in (5) is a higher order infinitesimal, we see that for values of $\underline{x}$ "sufficiently near" $\underline{a}$ that Δf is "approximately equal" to the portion of equation (5) that appears in braces, and this portion in braces is a linear combination of $x_1,\ldots,$ and Δx_n.

If this seems a bit "highbrow", let us observe that we had already used such results in Part 1 of this course when we studied functions of a single (independent) real variable. In that case, the geometric interpretation was that if a curve was smooth at a point, we could replace the curve by its tangent line at that point, provided that we remained "sufficiently close" to the point in question. In other words, the linear approximation given by (5) is a local property (as opposed to a global property), meaning that once we get far enough away from $\underline{a}$, the linear portion on the right hand side of (5) no longer is a reliable estimate of Δf.

In essence, then, if we study the local properties of functions of several variables, we may view the functions as being appropriate linear combinations of the variables, provided only that our functions are differentiable (so that equation (5) applies).

It is for this reason that the subject known as linear algebra (and in many respects this is a synonym for matrix algebra) finds its way into the modern treatment of functions of several variables. While we must keep in mind that there are other, independent reasons for studying matrix algebra, the fact that it has a "natural" application to functions of several real variables (which, after all, is the subject of this entire course) is enough reason to introduce the subject at this time.

It should be noted, however, that in any "game" of mathematics we never have to justify our reasons for making up definitions and rules. Consequently, while the present course material serves as motivation, the fact remains that our introduction to matrix algebra in this chapter can be understood without reference to functions of several real variables.

At any rate, our aim in this chapter is to introduce matrix algebra in its own right without reference (except for motivation) to the calculus of several variables. Then, once this is accomplished, we will, in the next chapter, revisit functions of several real variables from the new vantage point of matrix algebra.

B

An Introduction to Linear Combinations

Consider the two systems of equations (6) and (7) given below

$$\left.\begin{aligned} z_1 &= y_1 + 2y_2 - 4y_3 \\ z_2 &= 2y_1 - 3y_2 + 5y_3 \end{aligned}\right\} \qquad (6)$$

and

$$\left.\begin{aligned} y_1 &= x_1 + 3x_2 - 2x_3 + 4x_4 \\ y_2 &= 3x_1 - 2x_2 + 3x_3 - 5x_4 \\ y_3 &= 5x_1 + 8x_2 - 4x_3 + 2x_4 \end{aligned}\right\} \qquad (7)$$

In this case, it is easy (although, possibly tedious) to express z_1 and z_2 _explicitly_ in terms of x_1, x_2, x_3 and x_4. Namely, we simply replace y_1, y_2, and y_3 in (6) by their values given in (7), and we obtain

$$z_1 = (x_1 + 3x_2 - 2x_3 + 4x_4) + 2(3x_1 - 2x_2 + 3x_3 - 5x_4) - 4(5x_1 + 8x_2 - 4x_3 + 2x_4)$$

$$z_2 = 2(x_1 + 3x_2 - 2x_3 + 4x_4) - 3(3x_1 - 2x_2 + 3x_3 - 5x_4) + 5(5x_1 + 8x_2 - 4x_3 + 2x_4)$$

or

$$\left.\begin{aligned} z_1 &= -13x_1 - 33x_2 + 20x_3 - 14x_4 \\ z_2 &= 18x_1 + 52x_2 - 33x_3 + 33x_4 \, . \end{aligned}\right\} \qquad (8)$$

While obtaining (8) from (6) and (7) was not difficult, the fact is that by using specific coefficients, we may not have noticed a rather interesting relationship between how the coefficients in (6) and (7) were combined in order to obtain the coefficients in (8). To understand better what went on, we will restate equations (6) and (7) in a manner that better emphasizes the coefficients. This leads to the idea of double subscripts. For example, in equations (6), let us pick a single letter, say a, to represent a coefficient, and we shall then use a pair of subscripts, the first to tell us the row in which the coefficient appears and the other, the column. For example, by a_{13} we would mean the coefficient in the first row, third column on the right hand side of (6). Translated into the language of the z's and y's, this would mean the coefficient of y_3 in the expression for z_1. In terms of the actual coefficients given in equations (6), we have:

$a_{11} = 1,\ a_{12} = 2,\ a_{13} = -4$

$a_{21} = 2,\ a_{22} = -3,\ a_{23} = 5$.

In a similar way, we may generalize the coefficients in (7) by using the letter b with double subscripts. Again, in terms of our specific choice of coefficients in (7), we would have:

$b_{11} = 1,\ b_{12} = 3, b_{13} = -2,\ b_{14} = 4$

$b_{21} = 3,\ b_{22} = -2,\ b_{23} = 3,\ b_{24} = -5$

$b_{31} = 5,\ b_{32} = 8,\ b_{33} = -4,\ b_{34} = 2$

If we now rewrite equations (6) and (7) in this more general form, we obtain

$$\left.\begin{aligned} z_1 &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 \\ z_2 &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3 \end{aligned}\right\} \qquad (9)$$

and

$$\left.\begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 + b_{13}x_3 + b_{14}x_4 \\ y_2 &= b_{21}x_1 + b_{22}x_2 + b_{23}x_3 + b_{24}x_4 \\ y_3 &= b_{31}x_1 + b_{32}x_2 + b_{33}x_3 + b_{34}x_4 \end{aligned}\right\} \qquad (10)$$

Substituting the values for y as given in (10) into equations (9), we obtain

$$z_1 = a_{11}(b_{11}x_1 + b_{12}x_2 + b_{13}x_3 + b_{14}x_4) + a_{12}(b_{21}x_1 + b_{22}x_2 + b_{23}x_3 + b_{24}x_4) + a_{13}(b_{31}x_1 + b_{32}x_2 + b_{33}x_3 + b_{34}x_4)$$

$$z_2 = a_{21}(b_{11}x_1 + b_{12}x_2 + b_{13}x_3 + b_{14}x_4) + a_{22}(b_{21}x_1 + b_{22}x_2 + b_{23}x_3 + b_{24}x_4) + a_{23}(b_{31}x_1 + b_{32}x_2 + b_{33}x_3 + b_{34}x_4)$$

or

$$\begin{aligned} z_1 &= (a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31})x_1 + (a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32})x_2 + (a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33})x_3 + (a_{11}b_{14} + a_{12}b_{24} + a_{13}b_{34})x_4 \\ z_2 &= (a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31})x_1 + (a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32})x_2 + (a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{23})x_3 + (a_{21}b_{14} + a_{22}b_{24} + a_{23}b_{34})x_4. \end{aligned} \tag{11}$$

We may emphasize equations (11) further by introducing the notation

$$z_1 = c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + c_{14}x_4$$

$$z_2 = c_{21}x_1 + x_{22}x_2 + c_{23}x_3 + c_{24}x_4$$

where,

$$\left.\begin{array}{l|l} c_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ c_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & c_{22} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\ c_{13} = a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} & c_{23} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ c_{14} = a_{11}b_{14} + a_{12}b_{24} + a_{13}b_{34} & c_{24} = a_{21}b_{14} + a_{22}b_{24} + a_{23}b_{34} \end{array}\right\} \tag{12}$$

If we examine the subscripts in (12) we find the following relationship. Each c consists of a sum of three terms, each of the form ab. The <u>first</u> subscript of <u>a</u> matches the <u>first</u> subscript of <u>c</u>, the <u>second</u> subscript of <u>b</u> matches the <u>second</u> subscript of <u>c</u>, and the <u>second</u>

subscript of a matches the first subscript of b taking on all values 1,2,3. That is, using Σ-notation, equations (12) become

$$
\begin{array}{l|l}
c_{11} = \sum_{r=1}^{3} a_{1r}b_{r1} & c_{21} = \sum_{r=1}^{3} a_{2r}b_{r1} \\
c_{12} = \sum_{r=1}^{3} a_{1r}b_{r2} & c_{22} = \sum_{r=1}^{3} a_{2r}b_{r2} \\
c_{13} = \sum_{r=1}^{3} a_{1r}b_{r3} & c_{23} = \sum_{r=1}^{3} a_{2r}b_{r3} \\
c_{14} = \sum_{r=1}^{3} a_{1r}b_{r4} & c_{24} = \sum_{r=1}^{3} a_{2r}b_{r4}
\end{array}
\tag{13}
$$

Equations (13) can be further abbreviated by letting c_{ij} denote the coefficient of x_j in the expression for z_i. That is , we may abbreviate (13) as

$$c_{ij} = \sum_{r=1}^{3} a_{ir}b_{rj} \text{ where } i = 1,2 \text{ and } j = 1,2,3.$$

With this discussion in mind, we can generalize equations (6) and (7) as follows.

Let

$$
\left.
\begin{array}{l}
z_1 = a_{11}y_1 + \ldots + a_{1k}y_k \\
\vdots \\
z_n = a_{n1}y_1 + \ldots + a_{nk}y_k
\end{array}
\right\}
\tag{14}
$$

and let

$$
\left.
\begin{array}{l}
y_1 = b_{11}x_1 + \ldots + b_{1m}x_m \\
\vdots \\
y_k = b_{k1}x_1 + \ldots + b_{km}x_m
\end{array}
\right\}
\tag{15}
$$

then

$$
\begin{array}{l}
z_1 = c_{11}x_1 + \ldots + c_{1m}x_m \\
\vdots \\
z_n = c_{n}x_1 + \ldots + c_{nm}x_m
\end{array}
$$

where

$$c_{ij} = \sum_{r=1}^{k} a_{ir}b_{rj} \text{ ; } i = 1,\ldots,n \text{ and } j = 1,\ldots,m.$$

Notice that k,m, and n can be any positive integers [in equations (6) and (7), we had n = 2, k = 3, and m = 4] subject only to the conditions that the value of k in (14) must equal the value of k in (15). That is, $z_1,\ldots,$ and z_n are linear combinations of $y_1,\ldots,y_k$, and each of the variables $y_1,\ldots,y_k$ is a linear combination of $x_1,\ldots,x_n$.

This discussion is adequate to motivate the invention of matrix notation and the subsequent study of matrix algebra. We shall pursue this development in the next two sections.

C

An Introduction to Matrices

Matrix Algebra may be viewed as a game in the same way that many other mathematical systems have been introduced into our course as games. To say, however, that a matrix is a "rectangular array of numbers" (which is the common introductory definition) hardly describes (or justifies) the seriousness of the topic, nor the reasons that such great effort was expended to develop the subject.

For our purposes, there is no need to trace the chronological development of matrix algebra. Rather, it is sufficient to supply one meaningful motivation for the invention of matrix algebra in terms of the content of our present course.

Recall that we had considered systems of linear equations of the type,

$$\left.\begin{array}{l} z_1 = a_{11}y_1 + \ldots + a_{1k}y_k \\ \vdots \\ z_n = a_{n1}y_1 + \ldots + a_{nk}y_k \end{array}\right\} \qquad (16)$$

and

$$\left.\begin{array}{l} y_1 = b_{11}x_1 + \ldots + b_{1m}x_m \\ \vdots \\ y_k = b_{k1}x_1 + \ldots + b_{km}x_m \end{array}\right\} \qquad (17)$$

We then saw that we could express $z_1, \ldots,$ and z_n as linear combinations of $x_1, \ldots,$ and x_m, where the coefficients of the new linear system were completely determined by the coefficients of the equations (16) and (17).

In particular,

$$
\begin{array}{l}
z_1 = c_{11}x_1 + \cdots + c_{ij}x_j + \cdots + c_{1m}x_m \\
\quad \vdots \\
z_i = c_{i1}x_1 + \cdots + c_{ij}x_j + \cdots + c_{im}x_m \\
\quad \vdots \\
z_n = c_{n1}x_1 + \cdots + c_{nj}x_j + \cdots + c_{nm}x_m
\end{array}
$$

<u>where</u>

$$
\left.
\begin{aligned}
c_{ij} &= a_{i1}b_{ij} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj}; \quad i = 1,\ldots,n; \ j = 1,\ldots,m \\
&= \sum_{r=1}^{k} a_{ir}b_{rj} \quad \text{(if we use } \Sigma\text{-notation)}
\end{aligned}
\right\} \quad (18)
$$

Certainly, nothing in our development of equations (18) requires that we know anything about vector algebra, but it is interesting to note that whenever we are given an expression such as $x_1y_1 + \cdots + x_ny_n$, where the x's and the y's are real numbers, we may view the sum as the dot product of two n-tuples. Namely,

$$x_1y_1 + \cdots + x_ny_n = (x_1,\ldots,x_n)\cdot(y_1,\ldots,y_n)$$

[Recall that we used this technique in the textbook's approach to the gradient vector as a special directional derivative.]

In any event, if we elect to view (18) in terms of a dot product, we obtain

$$c_{ij} = (a_{i1},a_{i2},\ldots,a_{ik})\cdot(b_{1j},b_{2j},\ldots,b_{kj}). \quad (18a)$$

Our first vector on the right hand side of (18a), in terms of equations (16), represents the coefficients of $y_1,\ldots,$ and y_k in the expression for z_i. The second vector represents the different coefficients of the variable x_j in equations (17).

In other words, if we look at the arrays of coefficients in equations (16) and (17), the first factor on the right hand side of (18) seems to be a "row" vector, while the second vector seems to be a "column" vector. More suggestively, perhaps we should have written (18) as

$$c_{ij} = (a_{i1},\ldots,a_{ik})\cdot\begin{pmatrix} b_{ij} \\ b_{2j} \\ b_{kj} \end{pmatrix}. \tag{18}$$

Again, keep in mind that our discussion is in the form of foresight toward the invention of matrix algebra and that there is no great need to become ecstatic over equations (18a) and (18b). What is important is that we can now invent a convenient "shorthand" to summarize our results. Namely, we list the coefficients of equations (1) and (2) "in the order of their appearance". That is,

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} & \cdots & a_{1k} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{21} & a_{22} & \cdots & a_{2r} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ir} & \cdots & a_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nr} & \cdots & a_{nk} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{ij} & \cdots & b_{1m} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rj} & \cdots & b_{rm} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kj} & \cdots & b_{km} \end{pmatrix} \tag{19}$$

Each of the parenthesized expressions in (19) is then called a matrix. This, hopefully, supplies at least a partial motivation as to why a matrix is defined as a rectangular array of numbers.

To indicate more precisely the "size" of the rectangular array, we include the number of rows and columns which make up the matrix. (This is sometimes called the dimension of the matrix.) For example, if a matrix has 3 rows and 5 columns we refer to it as a 3 x 5 (or, 3 by 5) matrix, where it is conventional to list the number of rows first, followed by the number of columns. More generally, if p and q denote any positive whole numbers, by a p x q matrix we mean any rectangular array of numbers consisting of p rows and q columns.

It is also conventional to enclose the matrix in parentheses, just as we have done in (19). By way of a few examples,

$$\begin{pmatrix} 1 & 3 & -2 \\ 8 & 7 & 6 \end{pmatrix}$$

is called a 2 by 3 matrix since it has two rows and three columns.

$$\begin{pmatrix} \pi & \sqrt{2} & -7 & e & 2 \\ 8 & \sqrt{3} & 0 & \frac{1}{2} & 7 \\ 2 & 4 & \sqrt{7} & 8 & 0 \end{pmatrix}$$

is a 3 by 5 matrix since it has three rows and five columns. Notice from this example that the matrix is made up of <u>real numbers</u>, not necessarily integers.

Now it should be clear from our previous discussions in this course that newly-defined concepts <u>without structure</u> are of little, if any, benefit to us. In this sense, the expression given by (19) hardly helps us - until we compare it with equations (18), (18a), or (18b). Once we make this comparison it is not difficult to imagine a new matrix, which can be formed by suitably combining the two matrices in (19), which tells us how the coefficients look when $z_1, \ldots,$ and z_n are expressed in terms of $x_1, \ldots,$ and x_m.

In particular, the matrix we seek should be the one whose entry in the ith <u>row</u>, ith <u>column</u> is obtained by dotting the ith <u>row</u> of the <u>first</u> matrix in (19) [notice that each <u>row</u> of this matrix may be viewed as a k-tuple] with the ith <u>column</u> of the <u>second</u> matrix in (19) [again notice that each <u>column</u> of this matrix may be viewed as a k-tuple]. We are not saying that this procedure is "natural" what we are saying is that if we look at equation (18b) and the two matrices in (19), there is a <u>special</u> way of combining the two matrices in (19) to form a matrix which conveys the information required by equations (18).

We use this as motivation for defining the <u>product</u> of two matrices.

<u>Definition</u>

$$\left.\begin{aligned} &\begin{pmatrix} a_{11}\cdots a_{1k} \\ a_{n1}\cdots a_{nk} \end{pmatrix} \begin{pmatrix} b_{11}\cdots b_{1m} \\ b_{k1}\cdots b_{km} \end{pmatrix} = \begin{pmatrix} c_{11}\cdots c_{1m} \\ c_{n1}\cdots c_{nm} \end{pmatrix} \\ &\underline{\text{where}} \\ &c_{ij} = \sum_{r=1}^{k} a_{ir}b_{rj}, \text{ where } \begin{cases} i = 1, \ldots, n \\ j = 1, \ldots, m \end{cases} \end{aligned}\right\} \qquad (20)$$

or in n-tuple, dot product notation,

$$c_{ij} = (a_{i1},\ldots,a_{ik})\cdot\begin{pmatrix} b_{ij} \\ \vdots \\ b_{kj} \end{pmatrix} \quad ; \; i = 1,\ldots,n; \; j = 1,\ldots,m$$

where we have used the notion of a column vector (k-tuple) to emphasize what the ith <u>row</u> of the first matrix in (5) is being dotted with the jth <u>column</u> of the second matrix in (20).

<u>Note</u>:

1. All that is required in (20) to form the product of two matrices is that the number of <u>columns</u> in the <u>first</u> matrix equal the number of <u>rows</u> in the second <u>matrix</u>.

2. More specifically, definition (20) tells us that an n by k matrix multiplied by a k by m matrix is an n by m matrix (that is, the product gets its number of rows from the first matrix, and its number of columns from the second matrix).

3. An "easy" way to remember this alignment is in terms of equations (16) and (17). Namely, in equations (16) the number of <u>columns</u> (on the right hand side of the equations) is determined by $y_1,\ldots,$ and y_k; while in equations (17) the number of <u>rows</u> is determined by $y_1,\ldots,$ and y_k.

<u>Example</u>:

$$\begin{pmatrix} 2 & 3 & 4 & 5 \\ 3 & -1 & -2 & 6 \end{pmatrix}\begin{pmatrix} 3 & -1 & 2 \\ 1 & 2 & 3 \\ 1 & -1 & 1 \\ -2 & 1 & -1 \end{pmatrix}=\begin{pmatrix} 3 & 5 & 12 \\ -6 & 3 & -5 \end{pmatrix} \qquad (21)$$

Equation (21) follows from definition (20). For example to find the entry in the 2nd row, 3rd column of the matrix on the right hand side of (21), we dot the 2nd row of the first matrix on the left hand side of (21) with the 3rd column of the second matrix on the left hand side of (21). That is:

$$(3,\ -1,\ -2,\ 6)\cdot\begin{pmatrix} 2 \\ 3 \\ 1 \\ -1 \end{pmatrix} = (3,\ -1,\ -2,\ 6)\cdot(2,\ 3,\ 1,\ -1) = (3)(2) + (-1)(3) + (-2)(1) + (6)(-1) = -5$$

(Notice that commas are used when we write the row or column as a traditional n-tuple, but the commas are not used to separate entries of the matrix, or the components of a column vector.)

In (21) we saw that a 2 by 4 matrix multiplied by a 4 by 3 matrix yielded a 2 by 3 matrix, as should be the case.

To identify (21) with our motivation for "inventing" definition (20), notice that if

$$\begin{cases} z_1 = 2y_1 + 3y_2 + 4y_3 + 5y_4 \\ z_2 = 3y_1 - y_2 - 2y_3 + 3y_4 \end{cases}$$

and

$$\begin{cases} y_1 = 3x_1 - x_2 + 2x_3 \\ y_2 = x_1 + 2x_2 + 3x_3 \\ y_3 = x_1 - x_2 + x_3 \\ y_4 = -2x_1 + x_2 - x_3 \end{cases}$$

Then

$$\begin{cases} z_1 = 3x_1 + 5x_2 + 12x_3 \\ z_2 = -6x_1 + 3x_2 - 5x_3 \ . \end{cases}$$

As a final note for this section, let us keep in mind the fact that the definition of matrix multiplication may not seem natural, but by this time it is hoped that we understand the "game" sufficiently well so that we realize we make up rules and definitions in accordance with the problems we are trying to solve. In this respect, the idea of making a chain rule substitution into systems of linear equations is enough motivation for defining matrix multiplication as we did.

D

Matrix Algebra

Having defined matrices in the previous section, we now wish to define a structure (i.e., an algebra) on them. To this end, we first decide upon an equivalence relation by which we shall decide when two matrices may be said to be equal (equivalent). While our definition need not be based upon "reality", the fact is that our entire discussion has been motivated in terms of systems of linear equations. Thus, it seems reasonable that our definition of equivalent matrices should reflect this discussion.

Since we probably agree that two systems of equations must, among other things, have the same number of variables (dependent and independent), it seems realistic to require that two matrices have the same dimension (i.e., the number of rows in one must equal the number of rows in the other as must the number of columns) before we even consider them to be equivalent. In other words, if the number of rows are the same, then both sets of equations have the same number of dependent variables, while if the number of columns is the same, both sets have the same number of independent variables.

If we then agree that each equation is written in "lowest terms" (that is, the left hand side of each equation has a coefficient of 1), we see that the systems of equations are the same if and only if the two systems have the same coefficients, term by term.

With this as motivation, we begin our arithmetic of matrices by specifying a particular dimension and limiting our study (at any given time) to this particular dimension. Thus, we might begin with the set S of all m x n matrices, where m and n are fixed positive whole numbers (which may be equal, but don't have to be).

If a matrix belongs to our set S, we will denote it by a capital (upper case) letter (this is not at all crucial, but it fixes our notation). It is customary to use the corresponding lower case letter to denote the various entries of a matrix. Thus, a common notation might be that if $A \in S$ we will write A as (a_{ij}), where it is clear since i names the row and A has m rows, that i must equal either 1,2,..., or m. Similarly, since A has n columns and j denotes the column, j must equal one of the numbers 1,2,..., or n.

Sometimes we deal with more than one set of matrices at a time (that is, different dimensional matrices may be studied in the same investigation). For this reason, the set S is often written as S(m,n) to remind the "user" that we are considering m x n matrices specifically. When this notation is used, if we write $(a_{ij}) \varepsilon S(m,n)$, it is tacitly understood that $i = 1,\ldots,m$ and that $j = 1,\ldots,n$. To make sure that this notation is clear to you, simply observe that in S(2,3), for example, (a_{ij}) is an abbreviation for the 2 x 3 matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.$$

At any rate, let us now assume that we are dealing with the set of m x n matrices, S(m,n). We say that the two matrices $A = (a_{ij})$ and $B = (b_{ij})$, both of which are elements of S(m,n), are <u>equal</u> (<u>equivalent</u>), <u>written A = B</u>, if and only if

$$a_{ij} = b_{ij} \text{ for } \underline{\text{each}}\ i = 1,\ldots,m \text{ and } \underline{\text{each}}\ j = 1,\ldots,n. \qquad (23)$$

Notice that, while definition (22) was motivated by referring to our matrix "coding system" for handling equations (16) and (17), the definition stands on its own. Indeed, the verification that the definition of matrix equality as given by (22) is an equivalence relation is so elementary that we are even too embarassed to assign it as an exercise.

The next step in forming our abstract structure, with or without practical motivation, would be to define the "sum" of two matrices of S(m,n). Notice here, that "sum" means a <u>binary operation on elements of S(m,n)</u>, that is, a rule which tells us how to combine two elements of S(m,n) to form another element of S(m,n). According to this "loose" interpretation of "sum", it seems that the definition given by (20) for matrix multiplication could qualify as being called a "sum" since it combines matrices to form matrices. The problem is that definition (20) need not apply to elements of S(m,n)[that is, to m by n matrices]. For, among other things, we have seen that we can only multiply two matrices (according to "multiplication" as defined by (20) if the number of columns in the first matrix equals the number of rows in the second. In particular, then to multiply an m by n matrix by an m by n matrix, we must have that n = m (since the number of columns in the first matrix is n and the number of rows in the second is m). The point is, however, that we have agreed to pick m and n at random, so that it need not be true that $m = n$. (In fact, it seems to the contrary that if $m = n$, our choice was hardly random.)

To be sure, we could now make the restriction that $m = n$ (in which case an element of $S(m,n)$ is called a square matrix to indicate a rectangular array in which the number of rows is equal to the number of columns). For the time being, however, we would like to find a binary operation on $S(m,n)$ that does not require that $m = n$. One way of doing this is to agree to add two matrices term by term. That is, if (a_{ij}) and (b_{ij}) belong to $S(m,n)$, we will define the sum $(a_{ij}) + (b_{ij})$ to be the matrix (c_{ij}), where

$$(c_{ij}) = (a_{ij} + b_{ij}). \tag{23}$$

To illustrate definition (23), we have as an example:

$$\begin{pmatrix} 2 & 3 & 1 & 4 \\ -5 & 3 & 2 & -6 \end{pmatrix} + \begin{pmatrix} 3 & -2 & 3 & -1 \\ 7 & -4 & 3 & 5 \end{pmatrix} = \begin{pmatrix} (2+3) & (3-2) & (1+3) & (4-1) \\ (-5+7) & (3-4) & (2+3) & (-6+5) \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 1 & 4 & 3 \\ 2 & -1 & 5 & -1 \end{pmatrix}.$$

In this example, we have added two 2 by 4 matrices to obtain another 2 by 4 matrix.

An interesting facet of definition (23) is that it preserves our usual vector addition. Surprising as it may seem, we may view an m by n matrix as an mn-tuple. If this sounds like a tongue-twister, we are saying, for example, that a 2 by 3 matrix may be viewed as a 6-dimensional vector (6-tuple). For example, while it might be more convenient to write

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \tag{24}$$

than

$$(a_{11}, a_{12}, a_{13}, a_{21}, 2_{22}, a_{23}), \tag{25}$$

the fact is that our definition of matrix equality as given by equation (22) is the same as our definition of equality for vectors. With respect to the notations given by (24) and (25), to say that

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}$$

says that $a_{ij} = b_{ij}$ for each $i = 1,2$ and each $j = 1,2,3$; and this is the same as saying that

$$(a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}) = (b_{11}, b_{12}, b_{13}, b_{21}, b_{22}, b_{23}).$$

Of course once this analogy is made, it becomes rather natural to introduce the counterpart of scalar multiplication, and to this end, we have:

If $(a_{ij}) \varepsilon S(m,n)$ and c is any real number, we define

$$c(a_{ij}) = (ca_{ij}). \tag{26}$$

That is, definition (26) tells us that to multiply a matrix by a given scalar we multiply each entry of the matrix by the given scalar. This is analogous to scalar multiplication in vectors where to multiply a vector by a given scalar we multiplied each component of the vector by the given scalar.

At any rate, as we have said so often, we do not need to justify our definitions as given by (22), (23), and (26). We may simply accept them as rules of the game, or definitions. (We shall try to avoid the semantics of whether we are dealing with rules or with definitions. What is important is that they give us a mathematical structure on the set of matrices S(m,n) regardless of how we interpret them.)

Once (22), (23), and (26) have been accepted, we can, of course, prove various theorems about matrix addition and "scalar" multiplication. (Again, the proofs are trivial but possibly boring, so we shall not supply too many proofs here.)

For example,

(i) $A + b = B + A$.

For letting $A = (a_{ij})$ and $B = (b_{ij})$ we have

$$A + B = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}) = (b_{ij} + a_{ij})^* = (b_{ij}) + (a_{ij})$$

$$= B + A.$$

(ii) $(A + B) + C = A + (B + C)$.

$$\begin{aligned} \text{For, } [(a_{ij}) + (b_{ij})] + (c_{ij}) &= ([a_{ij} + b_{ij}]) + (c_{ij}) \\ &= ([a_{ij} + b_{ij}] + c_{ij}) \\ &= (a_{ij} + [b_{ij} + c_{ij}]) \\ &= (a_{ij}) + (b_{ij} + c_{ij}) \\ &= (a_{ij}) + [(b_{ij} + c_{ij})] . \end{aligned}$$

(iii) There exists $0 \varepsilon S(m,n)$ such that $A + 0 = A$ for each $A \varepsilon S(m,n)$.

For we need only define $0 = (a_{ij})$ by $a_{ij} = 0$ for each i and j. For example in $S(2,3)$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} a_{11} + 0 & a_{12} + 0 & a_{13} + 0 \\ a_{21} + 0 & a_{22} + 0 & a_{23} + 0 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

(iiii) Given $A \varepsilon S(m,n)$, there exists $-A \varepsilon S(m,n)$ such that $A + (-A) = 0$.

For if $A = (a_{ij})$, we need only let $-A = (-a_{ij})$. By way of illustration

---

* This is the crucial step, We know that a_{ij} and b_{ij} are <u>numbers</u> and for numbers, we already have the commutative rule for addition, i.e., $a_{ij} + b_{ij} = b_{ij} + a_{ij}$.

$$\begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix} + \begin{pmatrix} -2 & -3 & -4 \\ -5 & -6 & -7 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Other results are

(v) $c(A + B) = cA + cB$ where c is a scalar, A and $B \varepsilon S(m,n)$.

(vi) $c_1(c_2A) = (c_1c_2)A$ where c_1 and c_2 are scalars, and $A \varepsilon S(m,n)$.

(vii) $(c_1 + c_2)A = c_1A + c_2A$; c_1, c_2 scalars, $A \varepsilon S(m,n)$.

Notice that properties (i) through (vii) hold regardless of whether we are allowed to use the binary operation defined by definition (20), which we called the product of two matrices. If, however, we wish to be able to use this definition of a product, then, as we mentioned earlier, we may look at the special case $m = n$ for which the product as defined by (20) makes sense.

Thus, we now switch to the special case of $S(n,n)$ wherein we mean the square matrices of dimension n by n, for some fixed positive integer, n. Since properties (i) through (vii) hold for $S(m,n)$ regardless of the values of m and n, they hold, in particular, if $n = m$. In other words, our arithmetic on $S(n,n)$ inherits at the outset properties (i) through (vii) but in addition we may now talk about multiplication as defined by (20). We would now like to investigate to see what properties of matrix multiplication apply to our matrix algebra.

Again, rather than be too abstract at this stage of the development, let us omit formal proofs from here on, and instead demonstrate our results for the case $n = 2$, indicating, when feasible, how these results generalize. To begin with, if we use AB, as defined in (20), to denote the product of A and B, it turns out that we <u>cannot</u> conclude that $AB = BA$.

As an example, let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

Then

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 7 & 11 \end{pmatrix} \qquad (27)$$

while

$$BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 7 & 10 \end{pmatrix} \tag{28}$$

Since

$$\begin{pmatrix} 3 & 5 \\ 7 & 11 \end{pmatrix} \neq \begin{pmatrix} 4 & 6 \\ 7 & 10 \end{pmatrix}^{*}$$

a comparison of (27) and (28) shows that $AB \neq BA$.

The fact that AB need not equal BA does not mean that we cannot find elements of $S(n,n)$, A and B, such that $AB = BA$. For example, suppose now that we let

$$A = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{29}$$

While

$$BA = \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix} \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{30}$$

Comparing (29) with (30) we see that in this case $AB = BA$.

Granted that we do not need an "intuitive" reason as to why AB need not equal BA, the fact remains that it might be helpful if we could get a "feeling" about what is happening. To this end, it may prove helpful if we return to our interpretation in terms of systems of equations. For example, with A and B as in equations

* Recall our definition, that equality requires the matrices to be equal term by term. Thus, as soon as the term in the 1st row, 1st column of AB was unequal to the term in the 1st row, 1st column of BA, we could conclude that $AB \neq BA$.

(27) and (28), we see that AB denotes the systems of equations

$$\left.\begin{aligned} z_1 &= y_1 + 2y_2 \\ z_2 &= 3y_1 + 4y_2 \end{aligned}\right\} \text{ and } \left.\begin{aligned} y_1 &= x_1 + x_2 \\ y_2 &= x_1 + 2x_2 \end{aligned}\right\} \tag{31}$$

while BA denotes the systems

$$\left.\begin{aligned} z_1 &= y_1 + y_2 \\ z_2 &= y_1 + 2y_2 \end{aligned}\right\} \text{ and } \left.\begin{aligned} y_1 &= x_1 + 2x_2 \\ y_2 &= 3x_1 + 4x_2 \end{aligned}\right\} \tag{32}$$

Hopefully, it is clear that both systems (31) and (32) allow us to express z_1 and z_2 in terms of x_1 and x_2, but that it need not happen that both systems yield the same relationships between the z's and the x's.

What happened in equations (29) and (30) was a rather interesting special case. Without worrying here about how we picked this special example, look at the systems of equations defined by AB in (29). We have

$$\left.\begin{aligned} z_1 &= 9y_1 + 2y_2 \\ z_2 &= 4y_1 + y_2 \end{aligned}\right\} \text{ and } \left.\begin{aligned} y_1 &= x_1 - 2x_2 \\ y_2 &= -4x_1 + 9x_2 \end{aligned}\right\}. \tag{33}$$

Solving for z_1 and z_2 in terms of x_1 and x_2 in (33) yields

$$\left.\begin{aligned} z_1 &= 9(x_1 - 2x_2) + 2(-4x_1 + 9x_2) = x_1 \\ z_2 &= 4(x_1 - 2x_2) + (-4x_1 + 9x_2) = x_2 \end{aligned}\right\}. \tag{34}$$

From (34) we see that x_1 and x_2 in (33), may be replaced by z_1 and z_2, and this, in turn, says that equations (33) may be written as

$$\left.\begin{aligned} y_1 &= z_1 - 2z_2 \\ y_2 &= -4z_1 + 9z_2 \end{aligned}\right\} \tag{35a}$$

and

$$\left.\begin{aligned} y_1 &= z_1 - 2z_2 \\ y_2 &= -4z_1 + 9z_2 \end{aligned}\right\} . \qquad (36b)$$

Then, what we have shown in (34) is that the systems (35a) and (35b) are <u>inverses</u> of one another. That is, equations (35b) follow from (35a) merely by solving (35a) for y_1 and y_2 in terms of z_1 and z_2.

Conversely, equations (35a) may be obtained from equations (35b) by solving (35b) for z_1 and z_2 in terms of y_1 and y_2. In summary, in this example,

$$AB = BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

because if we substitute (35a) into (35b), or (35b) into (35a), we obtain either

$$\left.\begin{aligned} z_1 &= z_1 \\ z_2 &= z_2 \end{aligned}\right\} \quad \text{or} \quad \left.\begin{aligned} y_1 &= y_1 \\ y_2 &= y_2 \end{aligned}\right\} . \qquad (36)$$

A glance at (36) may provide a hint as to why $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is called the <u>identity</u> matrix (and is usually denoted by I*). That is,

$$\left.\begin{aligned} z_1 &= 1z_1 + 0z_2 \\ z_2 &= 0z_1 + 1z_2 \end{aligned}\right\} \text{, etc.}$$

Another reason that I_n is called the identity matrix follows from the

*In this special case, n = 2. In general one uses I_n to denote <u>the</u> element of S(n,n) each of whose diagonal entries is 1 and all other entries are 0. (By the diagonal entries of (a_{ij}) we mean $a_{11}, a_{22}, \ldots, a_{nn}$.) In other words, the identity matrix of S(n,n) is denoted by I_n and is the matrix

$$\begin{pmatrix} 1 & 0 & 0 \ldots 0 \\ 0 & 1 & 0 \ldots 0 \\ & & \\ 0 & 0 & 0 \ldots 1 \end{pmatrix}$$

fact that for each A S(n,n),

$$AI = IA = A, \tag{37}$$

[where, of course, the multiplication indicated in (37) is as given by (20).]

Again, in the special case n = 2, it is easy to see that

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Moreover, the structure behind why this happens is relatively easy to see for each choice of n. For example, the term in the ith row, jth column of $(a_{ij})I_n$ comes from dotting the ith row of (a_{ij}) with the jth column of I_n. But the jth column of I_n has <u>zeros</u> <u>everywhere</u> <u>except</u> in the jth row (by how I_n is defined). Thus, the required term is

$$a_{i1}0 + a_{i2}0 + \ldots + a_{ij}(1) + \ldots + a_{in}0 = a_{ij}.$$

While we have just seen that for two matrices A and B, it need not be true that AB = BA, there are properties of matrix multiplication that resemble properties of "regular" multiplication. We shall leave all proofs for the exercises, but it is not hard to show that if A,B, and C belong to S(n,n), then:

$$(AB)C = A(BC) \tag{38}$$

and

$$A(B + C) = AB + AC. \tag{39}$$

We have also seen that the matrix (a_{ij}) for which $a_{ij} = 0$ if $i \neq j$ and $a_{ij} = 1$ if $i = j$ is the identity matrix, I_n, and has the property that for all A in S(n,n),

$$AI_n = I_nA = A \quad (40)$$

Equations (38), (39), and (40), especially if we identify the matrix I_n with the number 1, are structurally equivalent to rules of regular arithmetic.

There is, however, one very major difference between matrix multiplication and numerical multiplication which we shall mention here, but the discussion of which will be postponed to the next section. Namely, in numerical arithmetic if $a \neq 0$, then there is a number denoted by a^{-1}, such that $a(a^{-1}) = 1$. In matrix algebra, it need not be true that for a non-zero matrix A (recall that non-zero means that at least one entry of the matrix is different from 0, since the zero matrix has 0's for each entry there exists a matrix denoted by A^{-1} such that

$$AA^{-1} = I_n \quad (41)$$

Notice that if our last assertion is true, it means that various theorems concerning numbers whose proofs depend on the fact that multiplicative inverses exist need not be true about matrices. In particular, there may be pairs of matrices, neither of which is the zero matrix, whose product is the zero matrix, or there may be matrices A,B, and C such that AB = AC but A is not the zero matrix and B does not equal C.

As illustrations, consider the facts that

$$\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 6 & 4 \\ -3 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (42)$$

although neither factor is the zero matrix, and

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 7 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 0 & 0 \end{pmatrix} \quad (43)$$

even though $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 3 & 4 \\ 7 & 2 \end{pmatrix} \neq \begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix}$

While from one point of view, it is with dismay that we do not have the luxury of the cancellation rule in matrix algebra, it turns out that all is not lost. Namely, in many of the most practical applications of matrix algebra, we are concerned only with matrices A for which A^{-1} exists. In fact such matrices are of such importance that they are given a special name - <u>non-singular matrices</u>. That is, A is called <u>non-singular</u> if A^{-1} exists, and it is called <u>singular</u> if A^{-1} does not exist.

The point is that for non-singular matrices, the cancellation law does apply. To see this, assume that A^{-1} exists and that $AB = AC$. We can then multiply both sides of the equality by A^{-1}, use the fact that multiplication is associative, and then conclude that $B = C$. In more detail,

$AB = AC \rightarrow$

$A^{-1}(AB) = A^{-1}(AC) \rightarrow$

$(A^{-1}A)B = (A^{-1}A)C \rightarrow$

$IB = IC \rightarrow$

$B = C$.

Similarly, if A is non-singular and $AB = 0$ then $B = 0$. Namely

$AB = 0 \rightarrow$

$A^{-1}(AB) = A^{-1}0 \rightarrow$

$(A^{-1}A)B = 0 \rightarrow$

$IB = 0 \rightarrow$

$B = 0$.

If we restrict certain statements to non-singular matrices, then, with the exception of the fact that multiplication is not commutative, the structure of "ordinary" algebra applies to matrix algebra.

In summary, let S_n denote the set of all n x n matrices. Then with equality, addition, and multiplication as defined previously; and with 0 and I also as defined before, we have:

1. $A + B = B + A$

2. $A + (B + C) = (A + B) + C$

3. $A + 0 = A$

4. If $A \varepsilon S_n$ then there exists $-A \varepsilon S_n$ such that $A + (-A) = 0$

5. $A(BC) = (AB)C$

6. $AI = IA = A$, for each $A \varepsilon S_n$*

7. $A(B + C) = AB + AC$

The rules which do not carry over from ordinary arithmetic are the commutative rule for multiplication and the rule of multiplicative inverses. In many pure mathematics studies we often single out matrices which commute with respect to multiplication (one such matrix, by definition is I), but in most practical cases it is crucial to assume that multiplication of matrices is a non-commutative operation.

On the other hand, the blow that multiplicative inverses need not exist is softened by the fact that we often deal with non-singular matrices so that the following two theorems are in effect:

(a) If A is non-singular and $AB = 0$, then $B = 0$.

(b) If A is non-singular and $AB = AC$, then $B = C$.

These two theorems need not be true if A is singular. (Certainly,

*Since it need not be true that $AB = BA$, it is not enough to say $AI = A$, for then since IA need not equal AI we could not conclude that IA also was equal to A. For that reason we state the rule as we did. In rule (3) we did not have to do this because by rule (1) we know that $A + 0 = 0 + A$. Therefore once $A + 0 = A$ it is a <u>theorem</u> that $0 + A = A$ since $0 + A = A + 0$.

it is possible that theorems (a) and (b) hold for certain matrices even if A is singular. For instance assume that A is any singular matrix and that both B and C are the zero matrix. Trivially, in this case, the results stated in the two theorems are true. What we mean in general is that if all we know is that A is singular and that B and C are arbitrary matrices, then we cannot conclude, without additional information, that if $AB = 0$, $B = 0$, or if $AB = AC$, that $B = C$.)

In the next section we shall focus our attention on non-singular matrices. In particular, we shall be interested in the analog of the algebraic equation

$ax = b$

which we solve in arithmetic as $x = a^{-1}b$ provided $a \neq 0$.*

The matrix analog will be that $AX = B$ and A is non-singular, then $X = A^{-1}B$.

The other computational problem that we shall investigate is that of trying to construct A^{-1} explicitly once A is a given non-singular matrix. In other words, while it is nice, for a given A, to know that A^{-1} exists, it is important that we be able to exhibit it if we hope to compute such expressions as $A^{-1}B$.

E

Matrix Equations

If A is an n x n matrix, we define A^{-1} to be that matrix such that $AA^{-1} = A^{-1}A = I_n$, where I_n is the n x n identity matrix.
Note that A^{-1} need not exist for a given matrix, A. The point is that if A^{-1} does exist, we may use the properties of matrix algebra

* If we want a further unifying thread between matrix algebra and numerical algebra, let us define a number to be non-singular if it has a multiplicative inverse. Then the "numerical" rule that if $a \neq 0$, a^{-1} exists, is equivalent to saying that 0 is the only singular number. In other words, the theorems that (i) $a \neq 0$ and $ab = 0$ imply that $b = 0$ and (ii) $a \neq 0$ and $ab = ac$ imply that $b = c$, may be restated as: (i') If a is non-singular then $ab = 0$ implies $b = 0$. (ii') If a is non-singular then $ab = ac$ implies $b = c$.

discussed in the previous section to solve the matrix equation*

$$AX = B. \tag{44}$$

In (44), we assume that A is a square matrix of dimension n x n, but that X can be any matrix of dimension n x m. That is, the product AX is defined as soon as the number of columns comes from the number of columns in the second factor.)

If A^{-1} exists, we solve (44) as follows

AX = B implies that

$$A^{-1}(AX) = A^{-1}B, \tag{45}$$

but, since matrix multiplication is associative,

$$A^{-1}(AX) = (A^{-1}A)X.$$

Putting this into (45), shows that

$$(A^{-1}A)X = A^{-1}B\ . \tag{46}$$

But by definition of A^{-1}, we have that $A^{-1}A = I_n$, so that (46) becomes

$$I_nX = A^{-1}B. \tag{47}$$

Finally, since I_n is the identity matrix, $I_nX = X$,** and, accordingly, (47) becomes

$$X = A^{-1}B. \tag{48}$$

---

* We do not talk about A^{-1} unless A is a square matrix. Among other reasons, we have the structural property that for both AA^{-1} and $A^{-1}A$ to be defined, it must happen that A is a square matrix. For if A is n x m, $AA^{-1} = I_n$ implies that A^{-1} is m x n. But if A^{-1} is m x n, $A^{-1}A$ cannot equal I_n unless m = n, since the number of rows in $A^{-1}A$ equals the number of rows in A^{-1}.

** The rule $I_nX = X$ as stated in the previous section required that X be an n x n matrix. Notice that as long as X is n by m (even if $m \neq n$) $I_nX = X$.

If we now compare (44) and (48) we see that a quick, but mechanical, way to solve $AX = B$ for X is to multiply both sides of the equation on the left* by A^{-1}. This mechanical approach is similar in structure to our approach in ordinary arithmetic where we solve $ax = b$ by dividing both sides of the equation by a (i.e., multiplying by a^{-1}) provided that $a \neq 0$.

Our major aim in this section is to show how we may determine A^{-1}, if it exists, once A is given. For once we can do this, the problem of solving matrix equations such as (44) becomes very easy. Before doing this, however, let us illustrate how the method of solving matrix equations works once we know A^{-1}. Recall that in equations (29) and (30) we saw that

$$\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix}\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

By definition of A^{-1}, this information tells us that if

$$A = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \text{ then } A^{-1} = \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix}.$$

Suppose now that X is the 2×3 matrix

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}.$$

Then B must also be a 2×3 matrix. For illustrative purposes let

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

* In numerical arithmetic, we do not have to worry about the order factors since $ab = ba$. In matrix algebra, multiplication need not be commutative. In particular, $A^{-1}B$ need not equal BA^{-1}, so order is important. In terms of (45), had we written

$AX = B$ implies $(AX)A^{-1} = BA^{-1}$

this would be permissible, but we could not "cancel" A and A^{-1} in $(AX)A^{-1}$ since this would require that $(AX)A^{-1} = A^{-1}(AX)$, which need not be true.

Under these conditions, equation (44) takes on the form

$$\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \tag{49}$$

By the way, notice that if we multiply the two matrices on the left side of (49) we obtain

$$\begin{pmatrix} 9x_{11} + 2x_{21} & 9x_{12} + 2x_{22} & 9x_{13} + 2x_{23} \\ 4x_{11} + x_{21} & 4x_{12} + x_{22} & 4x_{13} + x_{23} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

and, by definition of matrix equality, this yields the system of equations

$$\left.\begin{array}{l} 9x_{11} + 2x_{21} = 1 \\ 9x_{12} + 2x_{22} = 2 \\ 9x_{13} + 2x_{23} = 3 \\ 4x_{11} + x_{21} = 4 \\ 4x_{12} + x_{22} = 5 \\ 4x_{13} + x_{23} = 6 \end{array}\right\} \tag{50}$$

From a different point of view, then, equation (49) is a convenient shorthand notation for expressing equations (50).

At any rate, returning to (49) and mimicking our procedure in obtaining (48) from (44), we have

$$\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}^{-1} \left[\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{12} & x_{23} \end{pmatrix} \right] = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Therefore,

$$\Bigg[\underbrace{\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}}_{I_2} \Bigg] \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix} = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Therefore,

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_{11} & x_{12} & x_{13}\\ x_{21} & x_{22} & x_{23}\end{pmatrix} = \begin{pmatrix}9 & 2\\ 4 & 1\end{pmatrix}^{-1}\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}.$$

Therefore,

$$\begin{pmatrix}x_{11} & x_{12} & x_{13}\\ x_{21} & x_{22} & x_{23}\end{pmatrix} = \begin{pmatrix}9 & 2\\ 4 & 1\end{pmatrix}^{-1}\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}. \qquad (51)$$

In this particular example, we happen to know that

$$\begin{pmatrix}9 & 2\\ 4 & 1\end{pmatrix}^{-1} = \begin{pmatrix}1 & -2\\ -4 & 9\end{pmatrix}.$$

Hence, (51) becomes

$$\begin{aligned}\begin{pmatrix}x_{11} & x_{12} & x_{13}\\ x_{21} & x_{22} & x_{23}\end{pmatrix} &= \begin{pmatrix}1 & -2\\ -4 & 9\end{pmatrix}\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}\\ &= \begin{pmatrix}1-8 & 2-10 & 3-12\\ -4+36 & -8+45 & -12+54\end{pmatrix}\\ &= \begin{pmatrix}-7 & -8 & -9\\ 32 & 37 & 42\end{pmatrix}. \qquad (52)\end{aligned}$$

Equation (52) yields X explicitly.

Notice that by the definition of matrix equality, equation (52) also tells us that

$$\left.\begin{aligned}x_{11} &= -7\\ x_{12} &= -8\\ x_{13} &= -9\\ x_{21} &= 32\\ x_{22} &= 37\\ x_{23} &= 42\end{aligned}\right\}$$

and that this, in turn, is the solution of the system of equations (50).

Hopefully, this illustration starts to show us a connection between solutions of systems of <u>linear</u> equations and solutions of matrix equations, and, in particular, why matrix algebra has a natural application to systems of linear equations.

As a particular example, suppose we consider the system of m equations in n unknowns given by

$$\left.\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad \vdots \quad &\quad \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n &= b_m \end{aligned}\right\}. \tag{53}$$

Clearly, one does not need a knowledge of matrices to understand the system of linear equations given by (53). With the use of matrix notation, however, the system (53) has a very conveneient representation. Namely, we let A denote the m by n matrix (a_{ij}), we let X denote the n by 1 matrix (i.e., the column vector) whose entries are $x_1,\dots,x_n$; and we let B denote the m by 1 matrix whose entries are $b_1,\dots,b_m$. In other words, we rewrite (53) as

$$AX = B,$$

that is,

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} . \tag{54}$$

[As a check, multiplying the two matrices on the left side of (54) yields

$$\begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \tag{55}$$

and, recalling our definition of matrix equality, equation (55) is equivalent to the system (53)]

The key point is that to solve (53) it is sufficient to compute A^{-1}. For, in this event, equation (54) yields

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}^{-1} \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

whereupon we may equate $x_1,\ldots,x_n$ in terms of the a's and b's.

Again, by way of illustration, given the system of linear equations

$$\left.\begin{aligned} 9x_1 + 2x_2 &= 5 \\ 4x_1 + x_2 &= 1 \end{aligned}\right\} \tag{56}$$

we write

$$\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}.$$

Hence,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

but, since

$\begin{pmatrix} 9 & 2 \\ 4 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix}$, this means

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -4 & 9 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 5 - 2 \\ -20 + 9 \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ -11 \end{pmatrix}.$$

Hence, $x_1 = 3$, $x_2 = -11$.

[Check: $9(3) + 2(-11) = 5$, $4(3) + (-11) = 1$; so that $x_1 = 3$, $x_2 = -11$ is the solution of (56).]

We are not implying that matrix algebra should be used to solve two linear equations in two unknowns, but we are hoping that our simple examples are helping you feel more at home with inverse matrices and matrix equations. We also hope that you understand that arithmetical procedures for solving n linear equations in n unknowns become at best cumbersome for larger values of n. Consequently, a convenient device for finding A^{-1} may prove helpful in solving systems of linear equations.

At any rate, let us now introduce a technique for finding A^{-1} for a given A. There are many recipes for doing this, but we prefer to give a particularly simple interpretation (an interpretation that can be learned meaningfully even by the junior high school student). The interpretation we have in mind can, in fact, be presented independently of any knowledge of matrices.

By way of illustration, consider the system of equations

$$\left.\begin{aligned} x + 2y + 3z &= 7 \\ 2x + 5y + 8z &= 11 \\ 3x + 4y + 7z &= 14 \end{aligned}\right\} . \qquad (57)$$

We invoke the following facts without formal proof (hopefully they will seem "self-evident").

(i) If both sides of an equation are multiplied by a non-zero constant, the new equation has the same solution set as the old one. (For example, while $x + 2y + 3z = 7$ and $2x + 4y + 6z = 14$ are <u>different</u> equations, they are equivalent in the sense that a specific 3-tuple (x_0, y_0, z_0) is a solution of one equation <u>if and only if</u> it is also a solution of the other equation.)

(ii) If we replace any equation in a system by itself plus any other equation in the system, we again do not change the solution set of the original system. This result is often referred to "as equals added to equals are equal". (Again, by way of illustration, suppose we replace the first equation in (57) by the sum of the first and the third. This would give us a new system of equations

$$\left.\begin{array}{l} 4x + 6y + 10z = 21 \\ 2x + 5y + 8z = 11 \\ 3x + 4y + 7z = 14 \end{array}\right\} . \qquad (58)$$

The point is that (57) and (58), while being systems of different equations, still have the same solution set.)

Finally,

(iii) If we change the order in which the equations appear in a system, we again do not change the solution set of the system.

Together with these three "axioms", let us introduce the notation that when we say two systems of equations are <u>equivalent</u>, we mean that they have the same solution set. For example the systems (57) and (58) would be called equivalent; and let us agree to write, for example,

$$\left.\begin{array}{l} x + 2y + 3z = 7 \\ 2x + 5y + 8z = 11 \\ 3x + 4y + 7z = 14 \end{array}\right\} \sim \left.\begin{array}{l} 4x + 6y + 10z = 21 \\ 2x + 5y + 8z = 11 \\ 3x + 4y + 7z = 14 \end{array}\right\}$$

to indicate that two systems are equivalent.*

The point is that there is now an excellent system for solving equations such as (57), a system which we shall refer to as <u>the diagonalization process.</u>

Essentially, what we do is obtain an equivalent system of equations which has the first variable (unknown) appear nowhere below the first equation, the second unknown nowhere below the second equation, the third unknown nowhere below the third equation, etc. A modified version of this is to have the first unknown appear only in the first equation, the second unknown only in the second equation, etc.

*The interested reader should notice that $\sim$ as defined above is indeed an equivalence relation. That is (1) any system has the same solution set as itself, (2) if the first system has the same solution set as the second, then the second has the same solution set as the first, and (3) if the first and second systems have the same solution set and also the second and third systems, then the first system has the same solution set as the third.

If we use (57) as an example, we observe that if the second equation is replaced by itself minus twice the first equation, then the resulting equation has no x term in it. [In terms of our axioms (i), (ii), and (iii), we first replace the first equation by the equation we obtain when we multiply both sides by -2. In the <u>new</u> system, we then replace the second by the second plus the first. That is:

$$\begin{cases} x + 2y + 3z = 7 \\ 2x + 5y + 8z = 11 \\ 3x + 4y + 7z = 14 \end{cases} \sim \begin{cases} -2x - 4y - 6z = -14 \\ 2x + 5y + 8z = 11 \\ 3x + 4y + 7z = 14 \end{cases} \sim \begin{cases} -2x - 4y - 6z = -14 \\ y + 2z = -3 \\ 3x + 4y + 7z = 14 \end{cases} .$$

Replacing $-2x - 4y - 6z = -14$ by $x + 2y + 3z = 7$ yields

$$\begin{cases} x + 2y + 3z = 7 \\ 2x + 5y + 8z = 11 \\ 3x + 4y + 7z = 14 \end{cases} \sim \begin{cases} x + 2y + 3z = 7 \\ y + 2z = -3 \\ 3x + 4y + 7z = 14 \end{cases} .$$

In the last system of equations,we next replace the third equation by the third minus three times the first. In more detail,

$$\begin{cases} x + 2y + 3z = 7 \\ y + 2z = -3 \\ 3x + 4y + 7z = 14 \end{cases} \sim \begin{cases} -3x - 6y - 9z = -21 \\ y + 2z = -3 \\ 3x + 4y + 7z = 14 \end{cases} \sim \begin{cases} -3x - 6y - 9z = -21 \\ y + 2z = -3 \\ -2y - 2z = -7 \end{cases} .$$

Therefore, if we replace $-3x - 6y - 9z = -21$ by $x + 2y + 3z = 7$ and $-2y - 2z = -7$ by $2y + 2z = 7$, it follows that (57) is equivalent to

$$\begin{cases} x + 2y + 3z = 7 \\ y + 2z = -3 \\ 2y + 2z = 7 \end{cases} . \tag{59}$$

That is, the systems of equations given by (57) and (59) have the same solution set. The advantage of (59) lies in the fact that while it is "officially" a system of three equations in three unknowns, it is <u>effectively</u> a simpler system of two equations in two unknowns, since the last two equations in (59) involve only y and z; and once y and z are determined from these two equations, we find x immediately from the first equation. [Note: In practice we obtain (59) very quickly from (58), omitting several intermediate steps. For example, we usually say, given (59), "replace the 2nd equation by the 2nd minus twice the 1st, and replace the 3rd equation by the 3rd minus three times the 1st, whereupon (59) is obtained in one step.]

Our next step in the diagonalization process is to eliminate y (the 2nd variable) everywhere below the 2nd equation (in this case, in the 3rd equation). We do this by replacing the 3rd equation in (59) by the 3rd, minus twice the second. That is, (59) is equivalent to

$$\left\{\begin{aligned} x + 2y + 3z &= 7 \\ y + 2z &= -3 \\ -2z &= 13 \text{ [i.e., } 7-2(-3)] \end{aligned}\right. \tag{60}$$

We refer to the system (60) as being in diagonalized form. While (57) and (60) are equivalent, the beauty of equations (60) is that we never have to solve more than one equation in one unknown. For example, from the third equation in (60) we can immediately "pick off" the value for z (i.e., $z = -13/2$) then, with this value of z, we may go to the second equation in (60) to find the value of y, and then knowing both y and z, we can go to the first equation and determine x.

Another way of doing the same thing is to eliminate y from <u>every</u> equation <u>except</u> the 2nd in (60). To do this, we need only replace the first equation in (60) by the first minus twice the second. This would yield

$$\left.\begin{aligned} x \qquad + (-z) &= 13 \\ y + 2z &= -3 \\ -2z &= 13 \end{aligned}\right\} . \tag{61}$$

Again, while (61) is equivalent to (57), the advantage of (61) is that once we know z, we can find both x and y directly (independently of one another).

The final simplification of (61) occurs if we decide to eliminate z from both the first and second equations. There are several ways for doing this, but one rather straight-forward way is to replace the first equation in (61) by -2 times the first equation. This yields

$$\left.\begin{aligned} -2x \qquad + 2z &= -26 \\ y + 2z &= -3 \\ -2z &= 13 \end{aligned}\right\} \tag{62}$$

The coefficients in (62) are now "adjusted" so that we now need only to replace the first equation in (61) by the first plus the third, and the second equation by the second plus the third to obtain

$$\left.\begin{aligned} -2x \quad\quad\quad &= -13 \\ y \quad\quad &= 10 \\ -2z &= 13 \end{aligned}\right\} \tag{62'}$$

or

$$\left.\begin{aligned} x &= 13/2 \\ y &= 10 \\ z &= -13/2 \end{aligned}\right\} . \tag{63}$$

Again, while (57) and (63) are equivalent, notice that the solution set for (63) is particularly easy to write down by inspection!*

As a check, if we let $x = 13/2$, $y = 10$, and $z = -13/2$ in (57) we obtain

$$\begin{cases} \frac{13}{2} + 2(10) + 3(\frac{-13}{2}) = 20 - 13 = 7 \\ 2(\frac{13}{2}) + 5(10) + 8(\frac{-13}{2}) = 13 + 50 - 52 = 11 \\ 3(\frac{13}{2}) + 4(10) + 7(\frac{-13}{2}) = 40 + [\frac{39 - 91}{2}] = 40 - 26 = 14. \end{cases}$$

The technique in obtaining (63) from (57) did not depend on the specific values of the constants on the right hand side of the equations in (57). More generally, equations (57) could have been given in the form

$$\left.\begin{aligned} x + 2y + 3z &= b_1 \\ 2x + 5y + 8z &= b_2 \\ 3x + 4y + 7z &= b_3 \end{aligned}\right\} . \tag{64}$$

*Technically speaking the system

$$\left.\begin{aligned} x &= 13/2 \\ y &= 10 \\ z &= -13/2 \end{aligned}\right\}$$

is three equations in three unknowns. It is that the solution set $\{(13/2, 10, -13/2)\}$ of this system is "highly suggested" by the system itself.

Of course, it becomes complicated to keep track of the various computations when b_1, b_2, and b_3 are used in place of specific numbers. One device for handling (64) is the use of matrices as a <u>coding</u> or <u>place holder system</u>. That is, we may think of a 3 x 6 matrix in which the six columns are labeled x, y, z, b_1, b_2, and b_3, respectively.* In this way, for example, the row

2 4 3 8 5 6

would be an abbreviation for

$$2x + 4y + 3z = 8b_1 + 5b_2 + 6b_3.$$

Thus, the matrix code for (64) would be

$$\begin{array}{c} \begin{array}{ccc|ccc} x & y & z & b_1 & b_2 & b_3 \end{array} \\ \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 8 & 0 & 1 & 0 \\ 3 & 4 & 7 & 0 & 0 & 1 \end{array}\right) \end{array}. \tag{65}$$

We then perform the same operations on the matrix as we did upon the equations. That is, we replace the second row of (65) by the second row minus twice the first and the third row by the third row minus three times the first. This yields

$$\begin{array}{c} \begin{array}{ccc|ccc} x & y & z & b_1 & b_2 & b_3 \end{array} \\ \left(\begin{array}{cccc|cc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & -2 & -2 & -3 & 0 & 1 \end{array}\right)^{**} \end{array}. \tag{66}$$

For practice in understanding our matrix code, notice that (66) tells us that the system of equations

$$\begin{cases} x + 2y + 3z = b_1 \\ \quad\;\; y + 2z = -2b_1 + b_2 \\ \;\, - 2y - 2z = -3b_1 + b_3 \end{cases} \tag{67}$$

---

* More generally, had we been given n equations in n unknowns, our coding matrix would have been n by 2n with the columns "holding the place of" $x_1, \ldots, x_n, b_1, \ldots, b_n$.

**Analogous to our discussion about systems of equations, we call two matrices <u>(row)equivalent</u> (note, <u>not</u> equal) if one is obtained from the other by the three operations previously assumed for systems of equations.

is equivalent to the system (64).

Moreover, (67) actually tells us how it "evolved" from (64). Namely, observe that b_1, b_2, and b_3 identify each of the three equations in (64). That is b_1 refers to the first equation, b_2 to the second, and b_3 to the third. Thus, for example, $-2b_1 + b_2$ tells us to subtract twice the first equation from the second. In other words, the equation

$$y + 2z = -2b_1 + b_2$$

is obtained from (64) by substracting twice the first equation in (64) from the second equation in (64) (Thus, our matrix code always tells us how to check whether a new equation is correct.)

We next "reduce" (66) by making the entire second column (except for the 1 in the second row) consist of zeros. That is, we replace the first row of (66) by the first minus twice the second, and the third by the third plus twice the second. This yields

$$\begin{pmatrix} 1 & 0 & -1 & 5 & -2 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & 2 & -7 & 2 & 1 \end{pmatrix} . \tag{68}$$

Again, (68) "translates" into

$$\begin{cases} x \qquad - z = 5b_1 - 2b_2 \\ \quad y + 2z = -2b_1 + b_2 \\ \qquad 2z = -7b_1 + 2b_2 + b_3 \end{cases} \tag{69}$$

and (69) is equivalent to (64). Moreover, (69) tells us <u>at a glance</u> what <u>might not have appeared at all obvious in (64)</u>. For example, $2z = -7b_1 + 2b_2 + b_3$ tells us that if we add the third equation in (64) to twice the second and then subtract seven times the first, the x and y terms vanish. [As a check, $2(2x + 5y + 8z) + (3x + 4y + 7z) -7(x + 2y + 3z) = 2z$.]

Then, to complete the "diagonalization" of (68), we may replace the first row by its double to obtain

$$\begin{pmatrix} 2 & 0 & -2 & 10 & -4 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & 2 & -7 & 2 & 1 \end{pmatrix} \tag{70}$$

We then replace the first row in (70) by the sum of the first and third, and we replace the second row by the second minus the third, to obtain

$$\begin{pmatrix} 2 & 0 & 0 & 3 & -2 & 1 \\ 0 & 1 & 0 & 5 & -1 & -1 \\ 0 & 0 & 2 & -7 & 2 & 1 \end{pmatrix} . \tag{71}$$

Again, (71) tells us that (64) is equivalent to

$$\begin{cases} 2x = 3b_1 - 2b_2 + b_3 \\ y = 5b_1 - b_2 - b_3 \\ 2z = -7b_1 + 2b_2 + b_3 \end{cases} . \tag{72}$$

By way of further review, $y = 5b_1 - b_2 - b_3$ tells us that to solve for y in (64) we subtract the sum of the second and third equations from five times the first equation. As a check, $5(x + 2y + 3z) - (2x + 5y + 8z) - (3x + 4y + 7z) = y$, again a fact which might not seem so obvious when all we look at is the system (64).

The final step is to reduce the "first half" of matrix (71) to the identity matrix. This entails replacing the first row by one-half of itself, and the same thing applies to the third row. This yields

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 1 & 0 & 5 & -1 & -1 \\ 0 & 0 & 1 & -\frac{7}{2} & 1 & \frac{1}{2} \end{array}\right) . \tag{73}$$

Our matrix in (73) is a code for

$$\begin{cases} x = \frac{3}{2}b_1 - b_2 + \frac{1}{2}b_3 \\ y = 5b_1 - b_2 - b_3 \\ z = -\frac{7}{2}b_1 + b_2 + \frac{1}{2}b_3 \end{cases} , \tag{74}$$

The key point now is to observe that (74) is the <u>inverse</u> of (64) in the truest sense. That is, in (64) b_1, b_2, and b_3 are expressed as linear combinations of x, y, and z; while in (74) x, y, and z are expressed as linear combinations of b_1, b_2, and b_3. Our claim now is that our approach is going from (64) to (74) tells us how to invert a matrix.

In particular, if we let A denote the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 4 & 7 \end{pmatrix}$$

we then "<u>augment</u>" A by the matrix I_3 to form the 3 x 6 matrix

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 8 & 0 & 1 & 0 \\ 3 & 4 & 7 & 0 & 0 & 1 \end{array}\right) .$$

We then reduce this matrix as illustrated above, so that I_3 becomes the first half of the new matrix. That is

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 1 & 0 & 5 & -1 & -1 \\ 0 & 0 & 1 & -\frac{7}{2} & 1 & \frac{1}{2} \end{array}\right) .$$

Then the second half of the new matrix is the inverse of A.

In other words,

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 4 & 7 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ 5 & -1 & -1 \\ -\frac{7}{2} & 1 & \frac{1}{2} \end{pmatrix}.$$

As a check,

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 4 & 7 \end{pmatrix} \begin{pmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ 5 & -1 & -1 \\ -\frac{7}{2} & 1 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ 3 & -1 & -1 \\ -\frac{7}{8} & 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 4 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

To see why this is so, the point is that if we let

$$\begin{cases} w_1 = x + 2y + 3z \\ w_2 = 2x + 5y + 8z \\ w_3 = 3x + 4y + 7z \end{cases} \tag{75}$$

and

$$\begin{cases} x = \frac{3}{2}w_1 - w_2 + \frac{1}{2}w_3 \\ y = 5w_1 - w_2 - w_3 \\ z = -\frac{7}{2}w_1 + w_2 + \frac{1}{2}w_3 \end{cases} \tag{76}$$

then (75) and (76) are inverses of each other in the sense that if, for example, we substitute (76) into (75) we obtain

$$\begin{aligned} w_1 &= w_1 \\ w_2 &= w_2 \\ w_3 &= w_3 \end{aligned}$$

or

$$\left.\begin{aligned} w_1 &= 1w_1 + 0w_2 + 0w_3 \\ w_2 &= 0w_1 + 1w_2 + 0w_3 \\ w_3 &= 0w_1 + 0w_2 + 1w_3 \end{aligned}\right\}. \tag{77}$$

The matrix which represents the substitution of (76) into (75) is

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 4 & 7 \end{pmatrix}\begin{pmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ 5 & -1 & -1 \\ -\frac{7}{2} & 1 & \frac{1}{2} \end{pmatrix}$$

while the matrix which represents (77) is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, by what motivated our definition for matrix multiplication,

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 4 & 7 \end{pmatrix} \begin{pmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ 5 & -1 & -1 \\ -\frac{7}{2} & 1 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Additional drill is left for the exercises. In the remainder of this section, we want to show how our matrix coding system works in the event that A^{-1} doesn't exist. To this end, let the matrix A be given by

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}. \tag{78}$$

If we now set out to find A^{-1} by the method described earlier in this section, we form the 3 x 6 matrix

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{bmatrix} \tag{79}$$

and if we try to get I_3 as the first half of our <u>equivalent</u> matrix, we find that

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{bmatrix} \sim$$

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{bmatrix} \sim \tag{80}$$

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix} \tag{81}$$

where (81) is obtained from (80) by replacing the third row in (80) by the third minus twice the second.

If we recall our coding system, (79) represents the system of equations

$$\left.\begin{aligned} x + 2y + 3z &= b_1 \\ 4x + 5y + 6z &= b_2 \\ 7x + 8y + 9z &= b_3 \end{aligned}\right\} \tag{82}$$

while the third row of (81) tells us that

$$0\,[= 0x + 0y + 0z] = b_1 - 2b_2 + b_3 \,. \tag{83}$$

Now, (83) reveals some very interesting things to us. In the first place if b_1, b_2, and b_3 happen to be chosen in such a way that $b_1 - 2b_2 + b_3 \neq 0$, then the system of equations (82) has no solution. This means that we can not "invert" (82) to solve for x, y, and z as linear combinations of b_1, b_2, and b_3 (for if we could, this, by definition, would mean that (82) has a solution).

For example, if we let $b_1 = 1$, $b_2 = 3$, and $b_3 = 7$ (so that $b_1 - 2b_2 + b_3 \neq 0$), equation (82) becomes

$$\left.\begin{aligned} x + 2y + 3z &= 1 \\ 4x + 5y + 6z &= 3 \\ 7x + 8y + 9z &= 7 \end{aligned}\right\} \tag{82'}$$

whereupon if we subtract twice the second equation from the sum of the first and third, we obtain

$0 = 2\,[=b_1 - 2b_2 + b_3]$, which is an obvious contradiction, that is

if (82') had a solution, it would imply the absurd result that 0 = 2. For this reason a system of equations like (82') is called an incompatible system.

On the other hand if b_1, b_2, and b_3 are chosen so that $b_1 - 2b_2 + b_3 = 0$, the system of equations (82) has solutions. In fact, it then has too many solutions, for in this case, the relationship between b_1, b_2, and b_3 tells us that our equations are not independent (in fact, (83) tells us that the first equation minus twice the second plus the third must be identically zero), and, as a result, we effectively have three unknowns but only two (independent) equations. This means, for example, that we can pick one of our variables (unknowns) at random and then solve for the other two. That is, in this case, there is no unique way of expressing x, y, and z as linear combinations of b_1, b_2, and b_3. Again, by way of illustration, suppose $b_1 = 1$, $b_2 = 3$, $b_3 = 5$ (so that $b_1 - 2b_2 + b_3 = 0$). Then the equations

$$\left.\begin{aligned} x + 2y + 3z &= 1 \\ 4x + 5y + 6z &= 3 \\ 7x + 8y + 9z &= 5 \end{aligned}\right\} \qquad (82")$$

are compatible, but the third equation is redundant since it is equal to twice the second minus the first. In other words the system (82") is in effect two equations in three unknowns. Namely

$$\left.\begin{aligned} x + 2y + 3z &= 1 \\ 4x + 5y + 6z &= 3 \end{aligned}\right\}$$

and this system has infinitely many solutions, one for each specific choice of z. In other words, with b_1, b_2, and b_3 as in this illustration, there are infinitely many ways to express x, y, and z as linear combinations of b_1, b_2 and b_3. We shall pursue this idea in more computational detail in the exercises, but for now, we hope it is sufficiently clear as to why A^{-1} does not exist in this case.

In summary, given the n x n matrix $A = (a_{ij})$, A^{-1} exists if and only if the system of linear equations

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ \\ a_{n1}x_1 + \dots + a_{nn}x_n = b_n \end{cases}$$

allows us to express $x_1,\dots,x_n$ in a unique way as linear combinations of $b_1,\dots,b_n$, for all possible values of $b_1,\dots,b_n$.

Our matrix coding system tells us how to compute A^{-1} if it exists, and if it doesn't exist, our code tells us how the equations are dependent on one another, so that the system of equations for specific values of $b_1,\ldots,b_n$ either has no solution or else it has too many.

For those of us who have studied <u>determinants</u> (and this is mentioned only as an aside here since determinants will be studied in a self-contained manner as part of Block 7), we may recall that the existence of A^{-1} is equivalent to the statement that the determinant of A (written $|A|$ or det A) is <u>not</u> zero.

Before applying our study of linear algebra to the calculus of functions of several real variables we would like to "revisit" linear algebra from a more geometric point of view, and this shall be the discussion of the next section.

F

Linearity in Terms of Mappings

We have mentioned in our introduction to this chapter that any system of m equations in n unknowns may be viewed as a mapping from E^n into E^m. In particular, then, the linear system

$$\left.\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n \\ &\;\vdots \\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n \end{aligned}\right\} \tag{84}$$

may be viewed as that mapping of E^n into E^m which maps $\underline{x} = (x_1,\ldots,x_n)$ into $\underline{y} = (y_1,\ldots,y_m)$, where $y_1,\ldots,$ and y_m are as defined in (84).

Since the most interesting cases occur when n = m, we shall restrict the remainder of this section to the case in which we have n linear equations in n unknowns. The most familiar situation, one which we have studied quite exhaustively before, is when n = 1. In this case, the mapping may be viewed pictorially as a line through the origin in the xy-plane. That is, when n = 1, equations (84) reduce to the single equation, $y_1 = a_{11}x_1$, which, since there is but one independent and one dependent variable, we write as y = ax; and this graphs as a straight line of slope equal to <u>a</u>, passing through the origin.

The case $n = 2$ lends itself to a graphical interpretation, but it is not nearly as convenient as the case $n = 1$. Nevertheless, since a pictorial interpretation exists for $n = 2$, we shall discuss linearity in detail for this case, and simply generalize our results to higher dimensions where pictures fail us.

With $n = 2$, the system of equations (84) becomes

$$\left.\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned}\right\} . \tag{85}$$

Again, since there are only two independent variables and since we usually think of 2-space in terms of the xy-plane, it is conventional to rewrite (85) without subscripts, as

$$\left.\begin{aligned} u &= ax + by \\ v &= cx + dy \end{aligned}\right\} \tag{86}$$

where a,b,c, and d are fixed numbers.

The geometric interpretation of (86) is that it defines a mapping from the xy-plane into the uv-plane, defined by

$$(x,y) \rightarrow (u,v) = (ax + by,\ cx + dy).$$

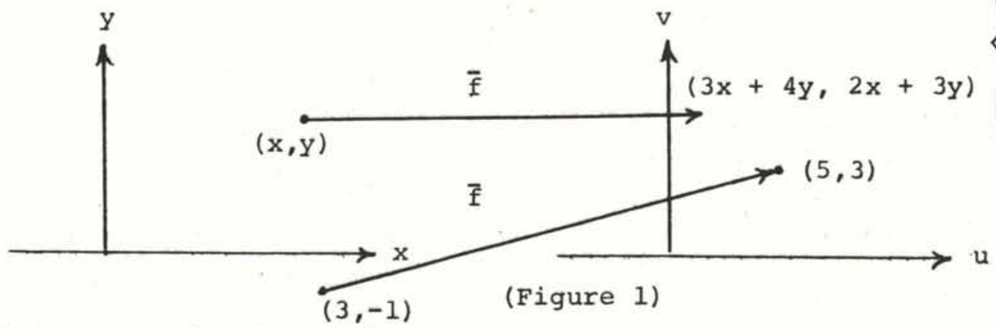
In order to keep our discussion as concrete as possible, let us, at least for time being, replace the abstract, general form of (86) by a particular example. Consider the mapping defined by

$$\left.\begin{aligned} u &= 3x + 4y \\ v &= 2x + 3y \end{aligned}\right\} \tag{87}$$

That is, the system (87) defines the mapping of the xy-plane into the uv-plane, defined by the fact that (x,y) in the xy-plane is mapped into $(3x + 4y,\ 2x + 3y) = (u,v)$ in the uv-plane.

That is, we may interpret (87), without reference to any picture (hence its adaptability to higher dimensions), as the mapping of $\underline{f}: E^2 \rightarrow E^2$ where for (x,y) dom $\underline{f}$, $\underline{f}(x,y) = (3x + 4y,\ 2x + 3y)$.

Pictorially, we have



i.e., $\underline{f}(3,-1) = [3(3)+4(-1), 2(3)+3(-1)] = (5,3)$

(Figure 1)

What properties does $\underline{f}$ have by virtue of its being linear that are not typical of non-linear functions?

From an analytic point of view, we have that

(1) $\underline{f}\ (\underline{a} + \underline{b}) = \underline{f}(\underline{g}) + \underline{f}(\underline{b})$

and

(2) If c is any scalar, $\underline{f}(c\underline{a}) = c\underline{f}(\underline{a})$.

These properties (which will be discussed in more detail analytically later in the course) have a particularly simple geometric interpretation that emphasizes the meaning of linear.

Namely, let us consider any straight line in the xy-plane which passes through the origin.

Except for the y-axis, all such lines have the form

$y = mx$

and in this case, if we replace y by mx in (87) we obtain

$$\left.\begin{aligned} u &= 3x + 4mx = x(3 + 4m) \\ v &= 2x + 3mx = x(2 + 3m) \end{aligned}\right\} \qquad (88)$$

so that

$$\frac{v}{u} = \frac{(2 + 3m)x}{(3 + 4m)x} = \frac{2 + 3m}{3 + 4m}$$

(unless $x = 0$*) so that

$$v = \left(\frac{2 + 3m}{3 + 4m}\right) u \qquad (88')$$

which in the uv-plane is the straight line through the origin, of slope

$$\frac{2 + 3m}{3 + 4m}\text{**}$$

In other words under a linear mapping $\underline{f}$: $E^2 \rightarrow E^2$ straight lines through the origin are mapped onto straight lines through the origin.

By way of illustration, we showed in Figure 1 that $\underline{f}(3,-1) = (5,3)$.

Now the line which connects (0,0) to (3,-1) has as its equation

$$\frac{y - 0}{x - 0} = \frac{-1 - 0}{3 - 0} \quad \text{or } y = -\frac{1}{3}x.$$

This line is the special case of (88) with $m = -\frac{1}{3}$. From (88') we see that this line is mapped into the line

$$v = \left[\frac{2 + 3(-\frac{1}{3})}{3 + 4(-\frac{1}{3})}\right] u, \quad \text{or} \quad v = \frac{3}{5}u.$$

Notice that (5,3) belongs to this line.

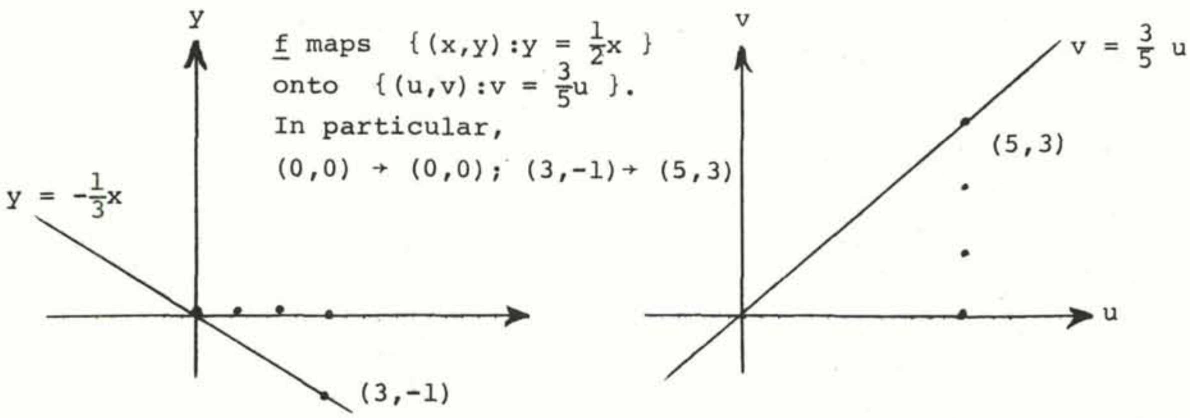
---

*$x = 0$ corresponds to the y-axis. In this case, (87) yields $u = 4y$, $v = 3y$, so that $v/u = 3/4$ (unless $y = 0$, but if both x and y equal 0, (87) simply says that $\underline{f}(0,0) = (0,0)$). That is, the mapping defined by (87) maps the y-axis onto the line $v = 3/4u$. Otherwise the mapping carires $y = mx$ **onto**

$$v = \left(\frac{2 + 3m}{3 + 4m}\right) u.$$

**If m is chosen so that $3 + 4m = 0$ (i.e., $m = -3/4$), equation (88') is not well-defined because of division by 0. When $m = -3/4$, (88) becomes $u = 0$, $v = -1/4x$. Since x may take on any values, (88) does not define v but does specify that $u = 0$. In other words when $m = -3/4$, we have the fact that $\underline{f}$ maps the line $y = -3/4x$ onto the v-axis (i.e., $u = 0$).

Thus Figure 1 is but a "piece" of the more general situation.



(Figure 2)

(Notice that $\underline{f}$ does not "preserve direction", that is, the image of $y = mx$ does not have to have slope m)

An interesting aspect of (88') is whether two different m values can produce the same value of $2 + 3m/3 + 4m$, for if this is possible, then two different lines in the xy-plane can map into the same line in the uv-plane.

Looking at

$$\frac{2 + 3m_1}{3 + 4m_1} = \frac{2 + 3m_2}{3 + 4m_2}$$

we have

$$6 + 8m_2 + 9m_1 + 12m_1m_2 = 6 + 9m_2 + 8m_1 + 12m_1m_2 \quad \text{or} \quad m_1 = m_2 \ .$$

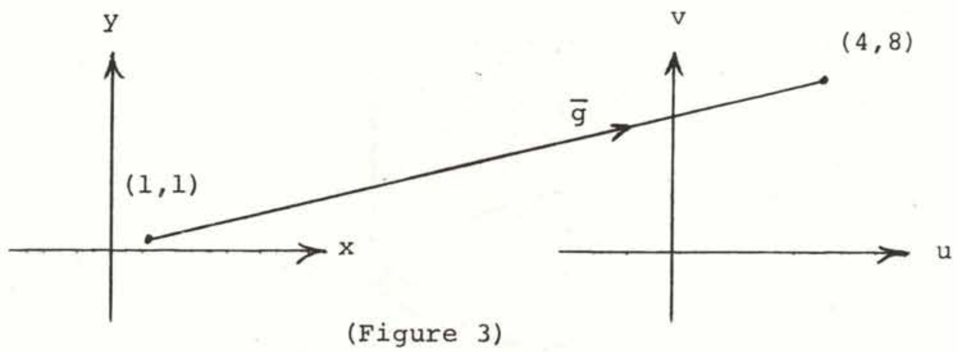
In different perspective, then, if $m_1 \neq m_2$, $y = m_1x$ and $y = m_2x$ are mapped into different lines in the uv-plane.

Let us, next, look at a different situation. Suppose we have

$$\left.\begin{aligned} u &= x + 3y \\ v &= 2x + 6y \end{aligned}\right\} \qquad (89)$$

System (89) defines the mapping $\underline{g}: E^2 \to E^2$ where $\underline{g}(x,y) = (x + 3y, 2x + 6y) = (u,v)$.

For example, $\underline{g}(1,1) = (1 + 3,\ 2 + 6) = (4,8)$. Pictorially,



(Figure 3)

We can again show that $\underline{g}$ maps lines passing through the origin into lines passing through the origin. In fact, if we let $y = mx$, we see from equations (89) that

$$\left.\begin{aligned} u &= x + 3mx = x(1 + 3m) \\ v &= 2x + 6mx = 2x(1 + 3m) \end{aligned}\right\} \tag{90}$$

From (90) we see that if $x \neq 0$ and $m \neq -\frac{1}{3}$

$$\frac{v}{u} = 2$$

or

$$\underline{v = 2u} \tag{91}$$

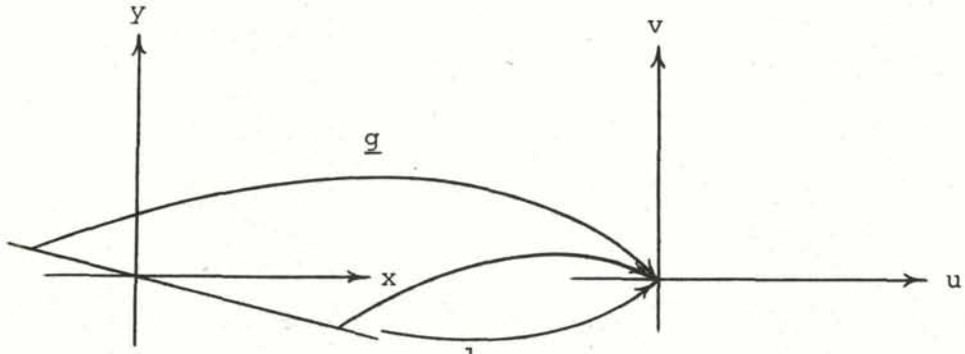
Equation (91) reveals the remarkable fact that with the possible exceptions of the y-axis ($x = 0$) and the line $y = -1/3x$ ($m = -1/3$), $\overline{g}$ maps every line through the origin into <u>the</u> line $v = 2u$ in the uv-plane.

In the exceptional cases, $x = 0$ and $m = -\frac{1}{3}$, we see from equations (89) that

(1) For $x = 0$; $u = 3y$ and $v = 6y$, so $v/u = 2$ again. Therefore, $\underline{g}$ maps the y-axis into the line $v = 2u$.

(2) For $m = -1/3$, equation (90) tells us that $u = v = 0$. Therefore, the line $y = -1/3x$ is mapped into $(0,0)$ by $\underline{g}$ (and clearly $(0,0)$ is on $v = 2u$). Therefore image $\underline{g} = \{(u,v):\ v = 2u\}$

Pictorially,



Every point on $y = -\frac{1}{3}x$ is mapped into (0,0).

(Figure 4)

What makes the line $y = -\frac{1}{3}x$ in the xy-plane and the point (0,0) in the uv-plane so important? Here again we see another interesting property of linearity. Let us look at any point on the line $v = 2u$, say $(u_o, 2u_o)$, where (0,0) is merely the special case, $u_o = 0$. Our first claim is that there is one and only one point on the y-axis that is mapped into $(u_o, 2u_o)$ by $\underline{g}$. Namely, if we return to equations (89) and let $u = u_o$, $v = 2u_o$, and $x = 0$ (to indicate that our point is on the y-axis), we see that

$$u_o = 3y$$

and equivalently

$$2u_o = 6y,$$

from which we conclude that $y = \frac{1}{3}u_o$.

In summary, the only point on the y-axis that is mapped into $(u_o, 2u_o)$ by $\underline{g}$ is $(0, \frac{1}{3}u_o)$. If we now shift the line $y = -\frac{1}{3}x$ to pass through this point (in other words, we look at the line $y = -\frac{1}{3}x + \frac{1}{3}u_o$), we find that every point in this line maps into $(u_o, 2u_o)$ as well! This may be verified directly from (89) by letting $y = -\frac{1}{3}x + \frac{1}{3}u_o$. Namely, we obtain

$$u = x + 3[-\frac{1}{3}x + \frac{1}{3}u_o]$$

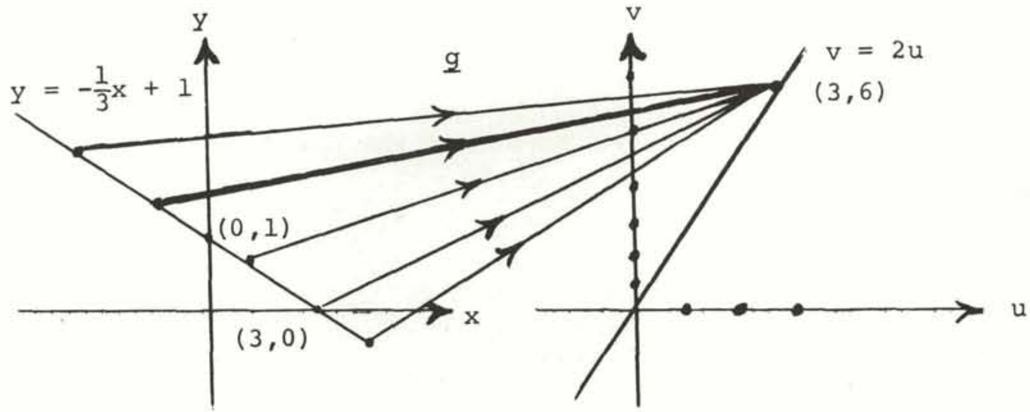
$$v = 2x + 6[-\frac{1}{3} + \frac{1}{3}u_o]$$

or

$$\left.\begin{array}{l} u = u_o \\ v = 2u_o \end{array}\right\} \quad \text{as required.}$$

As a concrete example, let us find the points in the xy-plane which are mapped onto the point (3,6) in the uv-plane. In the case, $u_o = 3$, so the line which maps into (3,6) is $y = -\frac{1}{3}x + 1$ (since $1 = \frac{1}{3}u_o$ in this case).

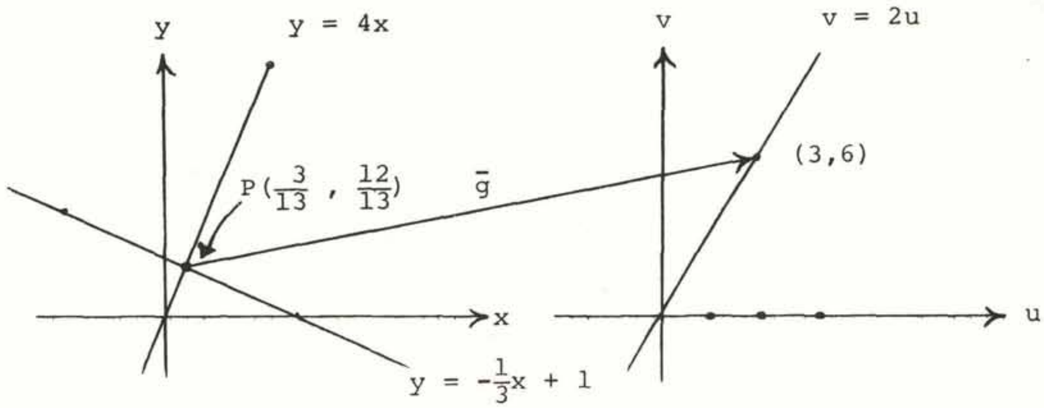
Pictorially



$\underline{g}(x,y) = (x + 3y, 2x + 6y)$

(Figure 5)

From Figure 5 we can construct the point on any line $y = mx$ that maps onto (3,6) under $\underline{g}$. Namely, we need only locate the point at which $y = -\frac{1}{3}x + 1$ intersects $y = mx$. For example, in Figure 6 we show how to locate the point (x,y) on $y = 4x$ such that $\underline{g}(x,y) = (3,6)$.



(Figure 6)

[Check: $x = \frac{3}{13}$, $y = \frac{12}{13}$ $\rightarrow$ $x + 3y = \frac{3}{13} + \frac{36}{13} = \frac{39}{13} = 3 = u$

$$2x + 6y = \frac{6}{13} + \frac{72}{13} = 6 = v$$]

In summary, the linear function $\underline{f}$ as defined by equation (87) maps E^2 onto E^2 in a 1-1 manner, while the mapping $\underline{g}$ defined by (89) is neither 1-1 nor onto.

What is it that distinguishes $\underline{f}$ and $\underline{g}$? Perhaps the best way to answer this problem is to look at the general case [equation (86)].

$$\left.\begin{aligned} u &= ax + by \\ v &= cx + dy \end{aligned}\right\}$$

Under this general situation we have that the mapping $\underline{h}:E^2 \rightarrow E^2$ defined by

$$\underline{h}(x,y) = (ax + by, cx + dy)$$

maps the line $y = mx$ into the line

$$\left.\begin{aligned} u &= ax + bmx \\ v &= cx + dmx \end{aligned}\right\}$$

or

$$v = [\frac{c + dm}{a + bm}]\ u \qquad (m \neq -\frac{a}{b}). \tag{92}$$

Let us see the conditions under which $y = m_1x$ and $y = m_2x$ can map into the same line in the uv-plane. According to (92) we would have

$$\frac{c + dm_1}{a + bm_1} = \frac{c + dm_2}{a + bm_2}$$

or

$$ac + bcm_2 + adm_1 + bdm_1m_2 = ac + adm_2 + bcm_1 + bdm_1m_2$$

or

$$(ad - bc)\ m_1 = (ad - bc)\ m_2 \quad . \tag{93}$$

From (93) we see at once that m_1 must equal m_2 provided that ad - bc $\neq 0$ (for if ad - bc = 0, (93) is automatically satisfied for all values of m_1 and m_2).

Referring to (86), ad - bc is precisely the determinant of coefficients. Utilizing the language of matrices, notice that (86) has the form

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and that this matrix equation has a unique solution for x and y in terms of u and v is and only if

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is non-singular. This is the same as saying that ad - bc $\neq 0$.

In summary (and we generalize to n-dimensional space without further discussion) then:

The linear mapping from E^n to E^n defined by the system of linear equations

$$\begin{aligned} y_1 &= a_{11}x_1 + \dots + a_{1n}x_n \\ &\vdots \\ y_n &= a_{n1}x_1 + \dots + a_{nn}x_n \end{aligned}$$

is 1-1 and onto if and only if the matrix of coefficients $A = [a_{ij}]$ is non-singular, i.e., if and only if A^{-1} exists (which means the determinant of A is unequal to 0).

Another way of saying this result in the case of our linear mappings of E^2 into E^2 is that under a linear mapping (0,0) is mapped into (0,0). That is, if u = ax + by and v = cx + dy, then clearly u and v are zero when x and y are zero. However, there may be other points in the (x,y) plane that map into (0,0). For linear mappings, it is enough to know that only (0,0) is mapped into (0,0) in order to conclude that the mapping is 1-1 and onto. In other words as soon as there is a point in the (x,y) plane, other than (0,0), that maps into (0,0), then the line determined by these two points will map into (0,0).

Thus, given the mapping $u = ax + by$, $v = cx + dy$ we look at $ad - bc$, and if this is not zero, we know that the mapping is 1-1 and onto. If it is zero, then the entire xy-plane is mapped into a single line (or in the most extreme case, the single point (0,0), that is, if $u = 0x + 0y$ and $v = 0x + 0y$). Namely, when $ad - bc = 0$, $c + dm$ is a constant multiple of $a + bm$ (details are left as an exercise) and we then have from (92) that $v = ku$. In this case the line $y = -\frac{a}{b}x$ maps into (0,0) and each line $y = -\frac{a}{b}x + u_o$ maps into a single point on the line $v = ku$.

At any rate, with all this as background, we end this chapter, and use the following chapter to apply our results to the calculus of functions of several real variables.