Unit 4: Linear Transformations

3.4.1(L)

Given any $v \varepsilon V = [u_1, u_2]$, we may write v uniquely in the form

$$v = x_1 u_1 + x_2 u_2. \tag{1}$$

Then, since f is linear,

$$f(v) = f(x_1 u_1 + x_2 u_2) \rightarrow$$
$$f(v) = x_1 f(u_1) + x_2 f(u_2). \tag{2}$$

We are told that

$$f(u_1) = -3u_1 + 2u_2 \tag{3}$$

and

$$f(u_2) = 4u_1 - u_2. \tag{4}$$

Hence, using (3) and (4) in (2), we conclude that

$$\begin{aligned} f(v) &= x_1(-3u_1 + 2u_2) + x_2(4u_1 - u_2) \\ &= (-3x_1 + 4x_2)u_1 + (2x_1 - x_2)\, u_2. \end{aligned} \tag{5}$$

If we agree to view $\{u_1, u_2\}$ as the coordinate system for V [i.e., we use (x_1, x_2) as an abbreviation for $x_1 u_1 + x_2 u_2$], we may rewrite (5) as

$$f(x_1, x_2) = (-3x_1 + 4x_2,\ 2x_1 - x_2). \tag{5'}$$

b. With $v_1 = 7u_1 + 5u_2$, we obtain from (5) that

$$\begin{aligned} f(v_1) &= (-21 + 20)u_1 + (14 - 5)u_2 \\ &= -u_1 + 9u_2; \end{aligned} \tag{6}$$

and with $v_2 = 2u_1 + 3u_2$, we see from (5) that

$$\begin{aligned} f(v_2) &= (-6 + 12)u_1 + (4 - 3)u_2 \\ &= 6u_1 + u_2. \end{aligned} \tag{7}$$

3.4.1(L) continued

Hence, from (6) and (7), we conclude that

$$f(v_1) + f(v_2) = 5u_1 + 10u_2 = 5(u_1 + 2u_2). \quad (8)$$

On the other hand,

$$\begin{aligned} v_1 + v_2 &= (7u_1 + 5u_2) + (2u_1 + 3u_2) \\ &= 9u_1 + 8u_2. \end{aligned}$$

Consequently, by (5), we see that

$$\begin{aligned} f(v_1 + v_2) &= (-27 + 32)u_1 + (18 - 8)u_2 \\ &= 5u_1 + 10u_2. \end{aligned} \quad (9)$$

Comparing (8) and (9), we conclude that $f(v_1 + v_2) = f(v_1) + f(v_2)$.

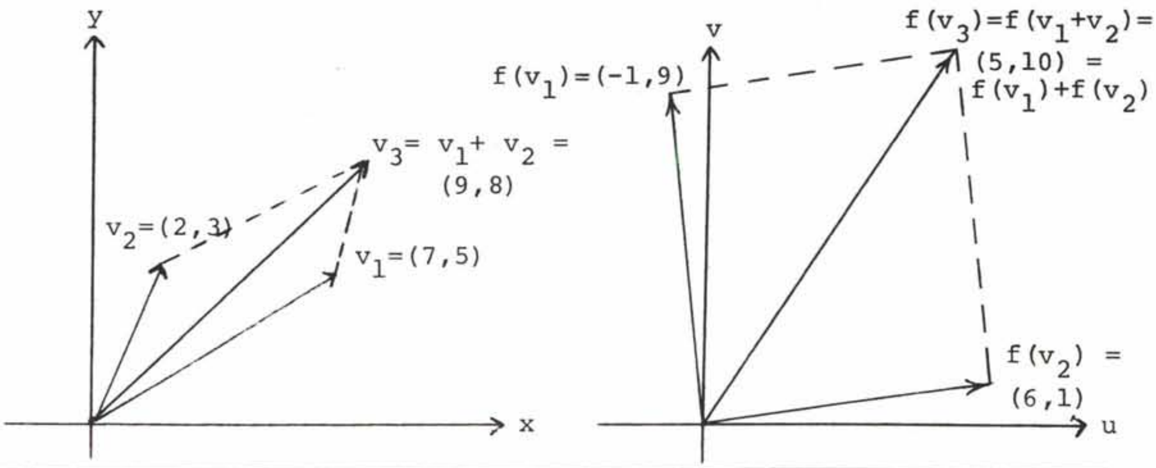
c. If we identify u_1 with $\vec{i}$ and u_2 with $\vec{j}$, we see from (5') that f is the linear mapping of the xy-plane into the uv-plane defined by

$$f(x,y) = (-3x + 4y, 2x - y). \quad (10)$$

That is,

$$\left.\begin{aligned} u &= -3x + 4y \\ v &= 2x - y \end{aligned}\right\} . \quad (10')$$

Pictorially,



3.4.1(L) continued

Notice from the figure that f "preserves structure"; that is, the image of $v_1 + v_2$ is the same as the sum of $f(v_1)$ and $f(v_2)$. Pictorially, parallelograms (with a vertex at the origin*) are mapped onto parallelograms.

Note:
We have limited our discussion in this exercise to the case n = 2 only so that our computations would be relatively easy and also so that we could still interpret our results pictorially. The point is that our results apply to any dimension. More specifically, if $V = [u_1, \ldots, u_n]$ and if $w_1, \ldots,$ and w_n are any n arbitrarily chosen elements in another space, W, then there is one and only one linear transformation $f: V \to W$ such that

$$\left.\begin{aligned} f(u_1) &= w_1 \\ f(u_2) &= w_2 \\ f(u_n) &= w_n \end{aligned}\right\} \qquad (1)$$

Namely, if $v = x_1u_1 + \ldots + x_nu_n$ is an arbitrary element of V, then f(v) must equal $x_1w_1 + \ldots + x_nw_n$. In different perspective, if we start with just the information given in (1) and then insist that f be linear; then the mapping defined by $f(v) = x_1w_1 + \ldots + x_nw_n$ "fills the bill" and no other linear mapping $f: V \to W$ can obey (1).

We must be careful in our reading of the above paragraph. We can't conclude that if dim V = n that we can define a linear mapping $f: V \to W$ just be requiring that any n elements in V be assigned to n elements in W. The key point is that the vectors in the domain must be a linearly independent set. That is, the property of linearity is such that $f(c_1v_1 + \ldots + c_nv_n)$ is determined once we know the effect of f on $v_1, \ldots,$ and v_n. Namely, by linearity,

$$f(c_1v_1 + \ldots + c_nv_n) = c_1f(v_1) + \ldots + c_nf(v_n).$$

*Identifying $x\vec{i} + y\vec{j}$ with (x,y) requires that $x\vec{i} + y\vec{j}$ originate at (0,0). Consequently, the natural tendency of thinking of a linear transformation as a mapping which maps lines into lines must be amended to say it maps lines <u>through the origin</u> into lines through the origin.

3.4.1(L) continued

For example, with reference to part (b) of the present exercise, once $f(v_1) = -u_1 + 9u_2$ and $f(v_2) = 6u_1 + u_2$, then f cannot be linear unless $f(v_1 + v_2) = 5u_1 + 10u_2$.

As a special case if dim V = n and m > n, then it is impossible to find a linear transformation $f:V \to W$ by arbitrarily prescribing the values of $f(v_1), \ldots,$ and $f(v_m)$. Namely, since m > n the set $v_1, \ldots, v_m$ is linearly dependent and this means that some of the v's are linear combinations of the others. In turn, if f is to be linear, this means that the images of these v's must be the same linear combination of the images of the other v's! In other words, if $v_k = c_1v_1 + \ldots + c_{k-1}v_{k-1}$, then $f(v_k) = c_1f(v_1) + \ldots + c_{k-1}f(v_{k-1})$. Hence, $f(v_k)$ cannot be arbitrarily prescribed once $f(v_1), \ldots,$ and $f(v_{k-1})$ have been chosen.

3.4.2

a. $V = [u_1, u_2, u_3]$.
$\alpha_1 = (1,2,3)$, $\alpha_2 = (4,5,6)$, $\alpha_3 = (7,8,9)$.

The major point is that since $\alpha_3 = 2\alpha_2 - \alpha_1$* and T is linear, then

$$\begin{aligned} T(\alpha_3) &= T(2\alpha_2 - \alpha_1) \\ &= 2T(\alpha_2) - T(\alpha_1). \end{aligned} \quad (1)$$

From (1) we see that $T(\alpha_3)$ is completely determined by the values of $T(\alpha_1)$ and $T(\alpha_2)$.

In particular, if $T(\alpha_1) = (3,1,2,4)$ and $T(\alpha_2) = (4,2,1,5)$, we see from (1) that

$$\begin{aligned} T(\alpha_3) &= 2(4,2,1,5) - (3,1,2,4) \\ &= (5,3,0,6). \end{aligned} \quad (2)$$

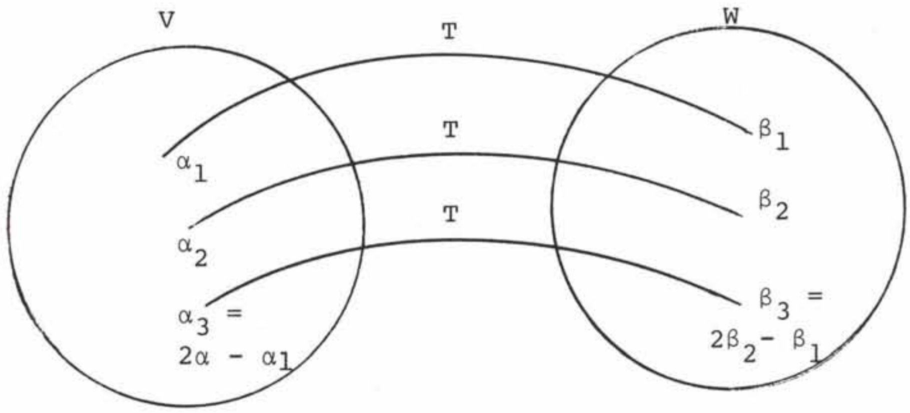
*This observation has nothing to do with linear transformations. Rather, it is obtained by the row-reduced matrix technique of the previous units wherein we discussed such things as finding the dimension of $S(\alpha_1, \alpha_2, \alpha_3)$.

3.4.2 continued

Thus, we see that if $T:V \to W$ is linear and $T(1,2,3) = (3,1,2,4)$ while $T(4,5,6) = (4,2,1,5)$, then $T(7,8,9)$ must equal $(5,3,0,6)$. Consequently, there is no linear transformation T from V to W such that

$$(1,2,3) \xrightarrow{T} (3,1,2,4)$$
$$(4,5,6) \xrightarrow{T} (4,2,1,5)$$
$$(7,8,9) \xrightarrow{T} (2,3,4,1).$$

Pictorially,



Analytically speaking, linear transformations "preserve" sums and scalar multiples. That is

$$v_1 \xrightarrow{T} w_1 \text{ and } v_2 \xrightarrow{T} w_2$$

implies that

$$a_1v_1 + a_2v_2 \xrightarrow{T} a_1w_1 + a_2w_2.$$

b. This is another form of (a). The only difference here is that since we have more vectors than dim V, the set must be linearly dependent. In particular, using our row-reduced matrix technique, we have

3.4.2 continued

$$\begin{array}{c} \begin{array}{ccccccc} u_1 & u_2 & u_3 & \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 0 & 1 & 0 & 0 \\ 2 & 3 & 5 & 0 & 0 & 1 & 0 \\ 3 & 7 & 6 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \sim \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -2 & 0 & 1 & 0 \\ 0 & 4 & 3 & -3 & 0 & 0 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & -5 & 1 & -4 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{c} \begin{array}{ccccccc} u_1 & u_2 & u_3 & \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & 1 & 3 & -2 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -4 & -9 & 5 & 1 \end{bmatrix} \end{array} \qquad (3)$$

The last row of (3) tells us that $-4\gamma_1 - 9\gamma_2 + 5\gamma_3 + \gamma_4 = 0$ or $\gamma_4 = 4\gamma_1 + 9\gamma_2 - 5\gamma_3$.

Consequently, $T(\gamma_4) = 4T(\gamma_1) + 9T(\gamma_2) - 5T(\gamma_3)$. Thus, we cannot prescribe $T(\gamma_4)$ arbitrarily once $T(\gamma_1)$, $T(\gamma_2)$, and $T(\gamma_3)$ are given.

3.4.3(L)

In the first two exercises we deliberately avoided any reference to matrices, at least with respect to linear transformations. The reason for this is that the concept of linear transformation does not require a knowledge of matrices. It is true, however, that judicious use of a matrix coding system allows us to express linear transformations rather nicely. We have made use of this matrix coding system not only in the present lecture but also in our introductory treatment of linear mappings in Block 4 of Part 2.

In the present exercise, what we want to do is emphasize how we may use matrix notation to compute f(v) for any v in V where $V = [u_1, \ldots, u_n]$. For the sake of simplicity we have assumed that V = W in this exercise but the general approach works for

3.4.3(L) continued

any linear transformation $f:V \to W$.

We also want to make sure that you are not confused by the two different conventions whereby linear transformations are represented by matrices. Consequently, we use the technique of the lecture in doing part (a), and the transpose of this method in doing part (b).

a. $V = [u_1,u_2]$

$f:V \to V$

$$f(u_1) = -3u_1 + 2u_2$$
$$f(u_2) = 4u_1 - u_2 \qquad (1)$$

Then, as we saw in Exercise 3.4.1(L), for any $v = x_1u_1 + x_2u_2 \varepsilon V$,

$$f(v) = (-3x_1 + 4x_2)u_1 + (2x_1 - x_2)u_2. \qquad (2)$$

If we now let

$$A = \begin{bmatrix} -3 & 4 \\ 2 & -1 \end{bmatrix}$$

be the transpose of the matrix of coefficients in (1) and if we let X denote the <u>column</u> vector (matrix)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

we have

$$AX = \begin{bmatrix} -3 & 4 & x_1 \\ 2 & -1 & x_2 \end{bmatrix}$$

$$\begin{bmatrix} -3x_1 + 4x_2 \\ 2x_1 - x_2 \end{bmatrix} \qquad (3)$$

Notice that (3) is a column vector whose components are those of f(v).

In summary, then, we may identify the matrix equation AX = Y with the vector equation (2), where

1. $A = \begin{bmatrix} -3 & 4 \\ 2 & -1 \end{bmatrix}$

3.4.3(L) continued

2. $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

3. $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

where $f(v) = (y_1, y_2) = y_1u_1 + y_2u_2$. Notice in this context that XA wouldn't make sense since X is 2 by 1 while A is 2 by 2.

b. If we want to use the matrix of coefficients in (1), then we let

$$B = \begin{bmatrix} -3 & 2 \\ 4 & -1 \end{bmatrix}^{*}.$$

In other words, $B = A^T$.

We now let X denote the <u>row</u> vector (matrix) $[x_1 \; x_2]$. In order not to confuse the row vector X here with the column vector X in part (a), let us agree to write $\vec{X}$ is S is a row matrix; and $\overset{\downarrow}{X}$ if X is a column vector.

At any rate, we easily verify that

$$\vec{X}B = [x_1 \quad x_2] \begin{bmatrix} -3 & 2 \\ 4 & -1 \end{bmatrix}$$

$$= [-3x_1 + 4x_2 \quad 2x_1 - x_2]$$

so that $\vec{X}B$ also names the components of f(v), but now as a row matrix.

The connection between parts (a) and (b) is simple. Namely, if

$$AX = Y \qquad (4)$$

Then

$$(AX)^T = Y^T.$$

*We use B rather than A only so that we do not confuse the notation in part (b) with that in part (a).

3.4.3(L) continued

Consequently,

$$X^T A^T = Y^T. \qquad (5)$$

If $X = \vec{X}$, then $X^T = \overset{\downarrow}{X}$; and conversely if $X = \overset{\downarrow}{X}$, $X^T = \vec{X}$

Thus, if we recall that $B = A^T$, we see from (4) and (5) that

$$A\overset{\downarrow}{X} = \overset{\downarrow}{Y} \rightarrow \vec{X}B = \vec{Y}$$

so that parts (a) and (b) are indeed two different matrix equations which code the same information.

It is not important which of the two conventions we use, but as we shall soon see, we must be careful not to confuse the two. For example, with A as in part (a), we could still compute $\vec{X}A$. We would obtain

$$[x_1 \quad x_2]\begin{bmatrix} -3 & 4 \\ 2 & -1 \end{bmatrix} = [-3x_1 + 2x_2 \quad 4x_1 - x_2]$$

which is a row vector <u>but it does not code the information given in (2)</u>. In other words, both $\vec{X}A$ and $A\overset{\downarrow}{X}$ make sense, but in general, the row vector defined by $\vec{X}A$ does not have the same components as the column vector AX.

The advantage of the notation used in (a) is that it allows us to write X after the matrix A and this preserves the notation f(x) as opposed to (x)f. The disadvantage is that we must remember that A is <u>not</u> the matrix of coefficients in (1) but rather the transpose of that matrix.

Conversely, the advantage of (6) is that B denotes the matrix of coefficients in (1); while the disadvantage is that we must write XA rather than AX.

3.4.4

$V = [u_1, u_2, u_3]$

$f: V \to V$ is linear and is defined by

$$\left.\begin{aligned} f(u_1) &= u_1 + u_2 + u_3 \\ f(u_2) &= 2u_1 + 3u_2 + 3u_3 \\ f(u_3) &= 3u_1 + 4u_2 + 6u_3 \end{aligned}\right\} \quad (1)$$

a. Letting

$$v = x_1u_1 + x_2u_2 + x_3u_3$$

we obtain from (1) that

$$\begin{aligned} f(v) &= x_1f(u_1) + x_2f(u_2) + x_3f(u_3) \\ &= x_1(u_1 + u_2 + u_3) + x_2(2u_1 + 3u_2 + 3u_3) \\ &\qquad + x_3(3u_1 + 4u_2 + 6u_3) \\ &= (x_1 + 2x_2 + 3x_3)u_1 + (x_1 + 3x_2 + 4x_3)u_2 \\ &\qquad + (x_1 + 3x_2 + 6x_3)u_3. \end{aligned} \quad (2)$$

b. Letting

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 3 \\ 3 & 4 & 6 \end{bmatrix} \text{ and } \vec{X} = [x_1\ x_2\ x_3],$$

we have

$$[x_1\ x_2\ x_3] \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 3 \\ 3 & 4 & 6 \end{bmatrix}$$

$$= [x_1 + 2x_2 + 3x_3 \quad x_1 + 3x_2 + 4x_3 \quad x_1 + 3x_2 + 6x_3].$$

In other words, if $f(v) = y_1u_1 + y_2u_2 + y_3u_3$, we may find y_1, y_2, and y_3 by letting $\vec{Y} = [y_1, y_2, y_3]$ and solving

$$\vec{X}B = Y. \quad (3)$$

c. Taking the transpose of both sides of (2), we obtain

$$(\vec{X}B)^T = \vec{Y}^T$$

3.4.4 continued

or

$B^T\vec{X}^T = \vec{Y}^T.$

Hence,

$$B^T\vec{X} = \vec{Y}. \quad (4)$$

Finally, if we let $A = B^T$, (4) becomes

$$A\vec{X} = \vec{Y}$$

where A is the transpose of the matrix of coefficients in (1).

3.4.5(L)

With $V = [u_1, u_2]$ and the linear mapping $f: V \to V$ defined by

$$\left.\begin{aligned} f(u_1) &= u_1 + 2u_2 \\ f(u_2) &= 3u_1 + 5u_2 \end{aligned}\right\} \quad (1)$$

it follows for any $v = x_1u_1 + x_2u_2 \varepsilon V$, that

$$\begin{aligned} f(v) &= x_1f(u_1) = x_2f(u_2) \\ &= x_1(u_1 + 2u_2) + x_2(3u_1 + 5u_2) \\ &= (x_1 + 3x_2)u_1 + (2x_1 + 5x_2)u_2 \end{aligned} \quad (2)$$

or in 2-tuple notation, relative to $\{u_1, u_2\}$ as a basis fo V,

$$f(x_1,x_2) = (x_1 + 3x_2, 2x_1 + 5x_2). \quad (3)$$

So far we have just been reviewing the technique discussed in the previous two exercises, but the new point we want to make is that while f does not depend on the choice of basis (i.e., the same linear transformation can be expressed in terms of [infinitely] many bases), the 2-tuple notation used in (3) does depend on the particular basis being used.

In other words, had we found another basis for V, say v_1 and v_2, and if $v\varepsilon V$; then f(v) would not depend on whether we wrote v in terms of

3.4.5(L) continued

terms of u_1 and u_2 coordinates or in terms of v_1 and v_2 coordinates, but the resulting 2-tuple notation would depend on the basis. This is what we want to show in part (a) of this exercise.

a. We are now told that we want to express f in terms of the vectors

$$v_1 = u_1 + u_2 \text{ and } v_2 = 2u_1 + u_2. \tag{4}$$

Again, primarily as a review, we may use row-reduction to show that v_1 and v_2 do indeed form a basis for V and at the same time we can see how v_1 and v_2 are expressed as linear combinations of u_1 and u_2. More specifically:

$$\begin{matrix} u_1 & u_2 & v_1 & v_2 \end{matrix}$$
$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1 \end{bmatrix}$$

so that

$$\left.\begin{aligned} u_1 &= -v_1 + v_2 \\ u_2 &= 2v_1 - v_2 \end{aligned}\right\} \tag{5}$$

Now we may compute $f(v_1)$ and $f(v_2)$ as follows: From (4),

$$\left.\begin{aligned} f(v_1) &= f(u_1 + u_2) \\ f(v_2) &= f(2u_1 + u_2) \end{aligned}\right\} \tag{6}$$

Then, by the linearity of f, we conclude from (6) that

$$\left.\begin{aligned} f(v_1) &= f(u_1) + f(u_2) \\ f(v_2) &= 2f(u_1) + f(u_2) \end{aligned}\right\} \tag{7}$$

so that by (1), (7) becomes

$$\left.\begin{aligned} &f(v_1) = (u_1 + 2u_2) + (3u_1 + 5u_2) = 4u_1 + 7u_2 \\ &\text{and} \\ &f(v_2) = 2(u_1 + 2u_2) + (3u_1 + 5u_2) = 5u_1 + 9u_2 \end{aligned}\right\} \tag{8}$$

3.4.5(L) continued

We may now use (5) to rewrite (8) as

$$\left.\begin{aligned} f(v_1) &= 4(-v_1 + v_2) + 7(2v_1 - v_2) = 10v_1 - 3v_2 \\ f(v_2) &= 5(-v_1 + v_2) + 9(2v_1 - v_2) = 13v_1 - 4v_2 \end{aligned}\right\},$$

from which we see that the matrix of coefficients of f relative to the bases $\{v_1, v_2\}$ is given by

$$B = \begin{bmatrix} 10 & -3 \\ 13 & -4 \end{bmatrix}. \tag{9}$$

b. Let

$$v = 4u_1 + 7u_2. \tag{10}$$

Then, $f(v) = y_1u_1 + y_2u_2$, where

$$[4 \quad 7] \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} = [y_1 \quad y_2]. \tag{11}$$

That is,

$$[4 + 21 \quad 8 + 35] = [y_1 \quad y_2].$$

Hence

$$y_1 = 25$$
$$y_2 = 43.$$

In other words,

$$f(v) = 25u_1 + 43u_2. \tag{12}$$

Notice that in terms of column vectors, (11) could have been rewritten as

$$\left([4 \quad 7] \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}\right)^T = [y_1 \quad y_2]^T,$$

or

3.4.5(L) continued

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

On the other hand, combining (5) and (10) we have

$$\begin{aligned} v &= 4(-v_1 + v_2) + 7(2v_1 - v_2) \\ &= 10v_1 - 3v_2. \end{aligned}$$

Hence, to compute $f(v) = z_1v_1 + z_2v_2$ we have

$$\begin{bmatrix} 10 & -3 \end{bmatrix} \begin{bmatrix} 10 & -3 \\ 13 & -4 \end{bmatrix} = \begin{bmatrix} z_1 & z_2 \end{bmatrix};$$

or

$$\begin{bmatrix} 100 - 39 & -30 + 12 \end{bmatrix} = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$$

so that

$$\begin{aligned} z_1 &= 61 \\ z_2 &= -18. \end{aligned}$$

That is, relative to the basis $\{v_1, v_2\}$

$$f(v) = 61v_1 - 18v_2. \qquad (13)$$

We may check (12) and (13) by replacing u_1 and u_2 in (12) by their values in (5). This yields

$$\begin{aligned} f(v) &= 25(-v_1 + v_2) + 43(2v_1 - v_2) \\ &= 61\ v_1 - 18v_2 \end{aligned}$$

which checks with (13).

From a geometric point of view, we are saying that a linear transformation of the xy-plane into the uv-plane is completely determined as soon as we know the image of a pair of non-parallel vectors in the xy-plane. In other words, a linear mapping is defined by its effect on one pair of non-parallel vectors, and there are infinitely many ways in which such a pair of vectors may be chosen. What does vary with the choice

3.4.5(L) continued

of basis, however, is the representation of a given vector as a linear combination of basis vectors.

What is interesting is how the matrix of coefficients,

$$\begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix},$$

in (1) is related to the matrix of coefficients,

$$\begin{bmatrix} 10 & -3 \\ 13 & -4 \end{bmatrix}$$

in (9). This is what we investigate in (c).

c. We have:

1. $V = [\alpha_1, \alpha_2]$.

2. $T(\alpha_1) = a_{11}\alpha_1 + a_{12}\alpha_2$
   $T(\alpha_2) = a_{21}\alpha_1 + a_{22}\alpha_2$

3. $\beta_1 = b_{11}\alpha_1 + b_{12}\alpha_2$
   $\beta_2 = b_{21}\alpha_1 + b_{22}\alpha_2$.

Assuming in (3) that $[\beta_1, \beta_2]$ is a basis for V, we may invert (3) to obtain

4. $\alpha_1 = c_{11}\beta_1 + c_{12}\beta_2$
   $\alpha_2 = c_{21}\beta_1 + c_{22}\beta_2$

where (3) and (4) are related by

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}^{-1} = C^{-1}.$$

If we now look at the <u>same</u> T but in terms of its effect on β_1 and β_2, we have by (3)

$$T(\beta_1) = T(b_{11}\alpha_1 + b_{12}\alpha_2)$$
$$T(\beta_2) = T(b_{21}\alpha_1 + b_{22}\alpha_2)$$

3.4.5(L) continued

so that by linearity

$$5.\quad \begin{aligned} T(\beta_1) &= b_{11}T(\alpha_1) + b_{12}T(\alpha_2) \\ T(\beta_2) &= b_{21}T(\alpha_1) + b_{22}T(\alpha_2). \end{aligned}$$

Using (2), (5) becomes

$$T(\beta_1) = b_{11}(a_{11}\alpha_1 + a_{12}\alpha_2) + b_{12}(a_{21}\alpha_1 + a_{22}\alpha_2)$$
$$T(\beta_2) = b_{21}(a_{11}\alpha_1 + a_{12}\alpha_2) + b_{22}(a_{21}\alpha_1 + a_{22}\alpha_2)$$

or

$$6.\quad \begin{aligned} T(\beta_1) &= (b_{11}a_{11} + b_{12}a_{21})\alpha_1 + (b_{11}a_{12} + b_{12}a_{22})\alpha_2 \\ T(\beta_2) &= (b_{21}a_{11} + b_{22}a_{21})\alpha_1 + (b_{21}a_{12} + b_{22}a_{22})\alpha_2. \end{aligned}$$

If we now replace α_1 and α_2 in (6) by their values in (4), we obtain

$$\begin{aligned} T(\beta_1) = {} & (b_{11}a_{11} + b_{12}a_{21})(c_{11}\beta_1 + c_{12}\beta_2) \\ & + (b_{11}a_{12} + b_{12}a_{22})(c_{21}\beta_1 + c_{22}\beta_2) \\ T(\beta_2) = {} & (b_{21}a_{11} + b_{22}a_{21})(c_{11}\beta_1 + c_{12}\beta_2) \\ & + (b_{21}a_{12} + b_{22}a_{22})(c_{21}\beta_1 + c_{22}\beta_2) \end{aligned}$$

or

$$7.\quad \begin{aligned} T(\beta_1) = {} & [(b_{11}a_{11} + b_{12}a_{21})c_{11} + (b_{11}a_{12} + b_{12}a_{22})c_{21}]\beta_1 \\ & + [(b_{11}a_{11} + b_{12}a_{21})c_{12} + (b_{11}a_{12} + b_{12}a_{22})c_{21}]\beta_2 \\ T(\beta_2) = {} & [(b_{21}a_{11} + b_{22}a_{21})c_{11} + (b_{21}a_{12} + b_{22}a_{22})c_{21}]\beta_1 \\ & + [(b_{21}a_{11} + b_{22}a_{21})c_{12} + (b_{21}a_{12} + b_{22}a_{22})c_{22}]\beta_2. \end{aligned}$$

It should be noted that our derivation of (7) was very straightforward, if a bit messy. However, if we think in terms of matrices, we see that the matrix of coefficients in (7) is very closely related to the matrices of coefficients in (2), (3), and (4). In particular, we have that:

3.4.5(L) continued

$$\begin{bmatrix} (b_{11}a_{11} + b_{12}a_{21})c_{11} + (b_{11}a_{12} + b_{12}a_{22})c_{21} & (b_{11}a_{11} + b_{12}a_{21})c_{12} + (b_{11}a_{12} + b_{12}a_{22})c_{21} \\ (b_{21}a_{11} + b_{22}a_{21})c_{11} + (b_{21}a_{12} + b_{22}a_{22})c_{21} & (b_{21}a_{11} + b_{22}a_{21})c_{12} + (b_{21}a_{12} + b_{22}a_{22})c_{22} \end{bmatrix}$$

$$= \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{12} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

$$= \left(\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right) \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

$= BAC$

$= BAB^{-1}.$

This result is of extreme importance. What it tells us is that if the n by n matrix A (we **have** dealt with the case $n = 2$ but this is of hardly any consequence) represents a linear transformation of V into V relative to the basis $\{\alpha_1, \ldots, \alpha_n\}$ then if B is any non-singular n by n matrix, we have that BAB^{-1} represents the same transformation relative to the basis $\{\beta_1, \ldots, \beta_n\}$ where

$$\beta_1 = b_{11}\alpha_1 + \ldots + b_{1n}\alpha_n$$
$$\beta_n = b_{n1}\alpha_1 + \ldots + b_{nn}\alpha_n \quad .$$

To see what this result means in terms of the present exercise, we form from part (a) that

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} .$$

3.4.5(L) continued

$$B = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}.$$

From B we see that B^{-1} is determined by

$$\left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{array}\right] \sim \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & -1 \end{bmatrix}$$

$$\left[\begin{array}{cc|cc} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1 \end{array}\right]$$

so that

$$B^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$$

Therefore,

$$B^{1}AB^{-1} = \left(\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \right) \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 7 \\ 5 & 9 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 10 & -3 \\ 13 & -4 \end{bmatrix}$$

which agrees with the matrix of coefficients in (9).

Note:
Two n by n matrices A and B are called similar if and only if there exists a non-singular matrix X such that $XAX^{-1} = B$. In terms of linear transformation of a vector space, the significance of similar matrices is that they represent the same linear transformation, but with respect to a (possibly) different basis.

From a structural point of view, it may be worth observing

3.4.5(L) continued

that the relation "is similar" is an equivalence relation. Namely, since $IAI^{-1} = A$, we have that any matrix A is similar to itself. Secondly if $XAX^{-1} = B$, then $X^{-1}(XAX^{-1})X = X^{-1}BX$, so that letting $Y = X^{-1}$, we have that $A = YBY^{-1}$. Hence, if A is similar to B then B is similar to A. Finally, if $XAX^{-1} = B$ and $YBY^{-1} = C$, we have by substitution that

$$Y(XAX^{-1})Y^{-1} = C$$

or

$$(YX)A(X^{-1}Y^{-1}) = C$$

or

$$(YX)A(YX)^{-1} = C$$

or

$ZAZ^{-1} = C$, where $Z = YX$.

Hence, if A is similar to B and B is similar to C, then A is similar to C.

What becomes confusing in the study of matrix algebra is the different types of equivalences involving matrices. For example, we have already talked about two types of equivalence; one when we talked about row-reducing matrices and again here in the context of linear transformations. It seems that one of the best ways not to get confused is to keep the model of a vector space in mind. In terms of such a model, we usually talk about row equivalence when we want to see whether two sets of vectors span the same space; and we talk about the equivalence defined by similar matrices (a similitude) when we want to examine whether two matrices represent the same linear transformation but perhaps with respect to a different basis.

In summary, similar matrices represent the same linear transformation but with respect to (possibly) different bases. More specifically, suppose

3.4.5(L) continued

$$V = [\alpha_1,\ldots,\alpha_n] = [\beta_1,\ldots,\beta_n]$$

and that

$$f(\alpha_1) = a_{11}\alpha_1 + \cdots + a_{1n}\alpha_n$$
$$\vdots$$
$$f(\alpha_n) = a_{n1}\alpha_1 + \cdots + a_{nn}\alpha_n$$

while

$$\beta_1 = b_{11}\alpha_1 + \cdots + b_{1n}\alpha_n$$
$$\vdots$$
$$\beta_n = b_{n1}\alpha_1 + \cdots + b_{nn}\alpha_n.$$

Then,

$$f(\beta_1) = c_{11}\beta_1 + \cdots + c_{1n}\beta_n$$
$$\vdots$$
$$f(\beta_n) = c_{n1}\beta_1 + \cdots + c_{nn}\beta_n$$

where the matrix $C = [c_{ij}]$ is obtained from the matrix $A = [a_{ij}]$ by $C = BAB^{-1}$ where $B = [b_{ij}]$.

As a final remark, notice that

$$C = BAB^{-1} \rightarrow$$
$$C^T = (BAB^{-1})^T \rightarrow$$
$$C^T = B^TA^T(B^{-1})^T \rightarrow$$
$$C^T = B^TA^T(B^T)^{-1}.$$

Hence, we see that the concept of "similar" does not depend on whether we deal with the matrices of coefficients or with their transposes.

3.4.6 (optional)

The major objective of this exercise is to offer an insight to the idea of what is meant by "invariant properties". In essence, this is part of the "Number-versus-Numeral" theme. Namely, we have a property which we wish to examine and this property does not depend on the "coordinate system" being used, but its appearance does. We must then make sure that when such a property is stated in terms of a particular coordinate system the derived consequences are valid in any coordinate system.

To apply this notion specifically to the ideas of this Unit, let us observe that the notion of a linear transformation defined on a vector space does not depend on the basis we are using for the vector space. The matrix of coefficients, however, does depend on the basis. Thus, we want to make sure that any consequences of a particular linear transformation are independent of the basis being used.

This investigation can become a bit sticky, so we have limited our investigation to two rather elementary situations. Namely, every vector space (other than the vector space which consists only of the zero vector) allows at least two linear transformations to be defined on it. In particular, there is the identity mapping defined by $f(v) = v$ for each $v \in V$; and the zero-mapping defined by $f(v) = 0$ for all $v \in V$. It is readily checked that both of these mappings are indeed linear.

a. Relative to a particular basis, the identity mapping is defined by the fact that each element of that basis is mapped into itself. Since this is true for any basis, it means that our matrix of coefficients in this case must be the n by n identity matrix I <u>no matter what basis is being considered</u>. In particular, then it means that according to our definition of similar matrices, no n by n matrix other than I would be similar to I.

Stated abstractly, we are saying that for the sake of consistency it had better turn out that if A is an n by n matrix and if X is also an n by n matrix such that $X^{-1}AX = I$, then $A = I$.

The proof of this result follows from the usual rules of matrix arithmetic.

Namely,

3.4.6 continued

$X^{-1}AX = I \rightarrow$
$X(X^{-1}AX)X^{-1} = X(I)X^{-1} \rightarrow$
$(XX^{-1})A(XX^{-1}) = (XI)X^{-1} \rightarrow$
$IAI = XX^{-1} \rightarrow$
$I(AI) = I \rightarrow$
$IA = I \rightarrow$
$A = I.$

Hence, as it should be, the only matrix similar to I is I itself.

b. If f is the 0-transformation and $\{\alpha_1,\ldots\alpha_n\}$ is any basis for V. Then $f(\alpha_i) = 0$ for $i = 1,\ldots,n$. Thus, relative to $\{\alpha_1,\ldots,\alpha_n\}$, the matrix of coefficients is 0 (the zero matrix) and this must be true for all bases of V. Thus, our definition of similar should also guarantee that the only matrix similar to the zero matrix is the zero matrix itself. To this end, then, suppose $X^{-1}AX = 0$. Then $X(X^{-1}AX)X^{-1} = X(0)X^{-1}$ or $(XX^{-1})A(XX^{-1}) = X(0X^{-1})$. Therefore, since $0A = 0$ for all matrices A, we have $IAI = X0$ or $I(AI) = 0$, whence $IA = 0$ or $A = 0$. That is, if A is similar to 0, then $A = 0$.

3.4.7

The matrix of f relative to $[u_1,u_2,u_3]$ is given by

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 3 \\ 3 & 5 & 4 \end{bmatrix}. \qquad (4)$$

We are also told that

$$\begin{aligned} v_1 &= u_1 + 2u_2 + 3u_3 \\ v_2 &= 2u_1 + 5u_2 + 6u_3 \\ v_3 &= 3u_1 + 6u_2 + 10u_3 \end{aligned} \qquad (2)$$

To show that $V = [v_1,v_2,v_3]$, we need only show that B^{-1} exists where B is the matrix of coefficients in (2). Using row-reduction again, we have

3.4.7 continued

$$\begin{array}{cccccc} u_1 & u_2 & u_3 & v_1 & v_2 & v_3 \end{array}$$

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 6 & 0 & 1 & 0 \\ 3 & 6 & 10 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 3 & 5 & -2 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{cccccc} u_1 & u_2 & u_3 & v_1 & v_2 & v_3 \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 14 & -2 & -3 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{bmatrix} \qquad (3)$$

From (3) we see that

$$B^{-1} = \begin{bmatrix} 14 & -2 & -3 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}. \qquad (4)$$

Therefore, $\{v_1, v_2, v_3\}$ is a basis for V and the matrix of coefficients for f relative to $\{v_1, v_2, v_3\}$ is given by

$$BAB^{-1} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 10 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 3 \\ 3 & 5 & 4 \end{bmatrix} \begin{bmatrix} 14 & -2 & -3 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \qquad (5)$$

$$\begin{bmatrix} 14 & 22 & 19 \\ 30 & 47 & 41 \\ 45 & 71 & 61 \end{bmatrix} \begin{bmatrix} 14 & -2 & -3 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \qquad (5')$$

$$\begin{bmatrix} 95 & -6 & -23 \\ 203 & -13 & -49 \\ 305 & -19 & -74 \end{bmatrix}. \qquad (5'')$$

"Decoding" (5"), we have

$$\left.\begin{array}{l} f(v_1) = 95v_1 + 50v_2 - 23v_3 \\ f(v_2) = 203v_1 - 13v_2 - 49v_3 \\ f(v_3) = 305v_1 - 19v_2 - 74v_3 \end{array}\right\} \qquad (6)$$

3.4.7 continued

Partial Check:

$$\begin{aligned} f(v_1 &= f(u_1 + 2u_2 + 3u_3) \\ &= f(u_1) + 2f(u_2) + 3f(u_3) \\ &= (u_1 + u_2 + u_3) + 2(2u_1 + 3u_2 + 3u_3) + 3(3u_1 + 5u_2 + 4u_3) \\ &= 14u_1 + 22u_2 + 19u_3 \end{aligned}$$

[and this checks with the first row of first matrix in (5')].

Hence by (3),

$$\begin{aligned} f(v_1) &= 14(14v_1 - 2v_2 - 3v_3) + 22(-2v_1 + v_2) + 19(-3v_1 + v_3) \\ &= 95v_1 - 6v_2 - 23v_3 \end{aligned}$$

[which checks with the first row of (5")].

3.4.8(L)

Some Basic Review:

Suppose that $f:V \to W$ is linear (where W may be the same as V but doesn't have to be); and let N_f denote the null space of V relative to f. That is,

$$N_f = \{v \varepsilon V: f(v) = 0\} . \qquad (1)$$

We have already seen in the lecture that N_f is indeed a subspace (rather than just a subset) of V and that N_f is not empty since at least the zero element of V belongs to N_f.

Some Additional Terminology

The image of V relative to f, that is, f(V), where

$$f(V) = \{f(v): v \varepsilon V\}$$

which we already know is a subspace of W is called the rank of f. Moreover by the nullity of f we mean dim N_f.

3.4.8(L) continued

Main Result

If dim V = n and $f:V \to W$ is linear, then dim V = dim N_f + dim R_f where R_f denotes the rank of f[i.e., f(V)].

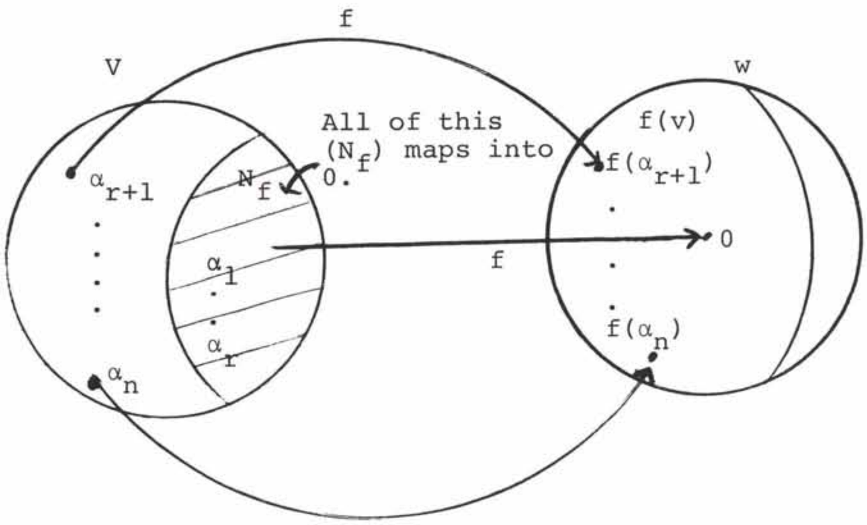
Interpretation of the Main Result

Suppose dim N_f = r. Clearly r may be as small as zero, and this will be the case if N_f consists of 0 alone; or as great as n, and this will be the case if N_f = V [i.e., if f(v) = 0 for every $v \varepsilon V$].

In any event, if we let $\{\alpha_1,\ldots,\alpha_r\}$ be a basis for N_f, we may augment this basis to become a basis for V. That is, we may find $\alpha_{r+1},\ldots,\alpha_n \varepsilon V$ such that $V = [\alpha_1,\ldots,{}_r,\alpha_{r+1},\ldots,\alpha_n]$. Our claim is that

$$f(v) = [f(\alpha_{r+1}),\ldots,f(\alpha_n)]. \qquad (1)$$

That is, N_f is mapped onto 0 and the image of f is determined by the images of $\alpha_{r+1},\ldots,$ and α_n. Pictorially,



A formal proof is given as an optional note at the end of this exercise but for now we prefer to illustrate the main ideas in terms of this explicit example.

We are given that

$$V = [u_1,u_2,u_3] \qquad (2)$$

3.4.8(L) continued

and that $f:V \to V$ is the linear transformation defined by

$$\begin{aligned} f(u_1) &= u_1 + 2u_2 + 3u_3 \\ f(u_2) &= 2u_1 + 5u_2 + 8u_3 \\ f(u_3) &= u_1 + 4u_2 + 7u_3 \end{aligned} \qquad (3)$$

We want to describe $f(v)$. What should be immediately clear is that $\{f(u_1), f(u_2), f(u_3)\}$ spans $f(v)$. The proof is clear since each $v \varepsilon V$ may be written in the form $v = c_1u_1 + c_2u_2 + c_3u_3$, whereupon

$$f(v) = f(c_1u_1 + c_2u_2 + c_3u_3). \qquad (4)$$

Hence, by the linearity of f, we see from (4) that

$$f(v) = c_1f(u_1) + c_2f(u_2) + c_3f(u_3), \qquad (5)$$

and since $f(v)$ is any member of $f(v)$, we deduce from (5) that every member of $f(v)$ may be expressed as a linear combination of $f(u_1)$, $f(u_2)$, and $f(u_3)$*.

What we can't say for sure, in general, is whether $\{f(u_1), f(u_2), f(u_3)\}$ is linearly independent. In our particular exercise we can use (3) together with our by-now-hopefully-familiar row-reduced matrix code to fine a basis for $f(v)$. That is, $f(v)$ is spanned by $\{f(u_1, f(u_2), f(u_3)\}$ and we obtain from (3):

*We say "may be expressed as..." since $\{u_1,u_2,u_3\}$ is but one particular basis for V. In this respect, there is nothing special about $\{u_1,u_2,u_3\}$ except that as the problem stands, it indicates the "coordinate system" being used to describe V.

3.4.8(L) continued

$$\begin{array}{cccccc} u_1 & u_2 & u_3 & f(u_1) & f(u_2) & f(u_3) \\ \hline \end{array}$$

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

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 2 & 4 & -1 & 0 & 1 \end{bmatrix} \sim$$

$$\begin{array}{l} \\ \beta_1 = \\ \beta_2 = \\ \\ \end{array} \begin{array}{cccccc} u_1 & u_2 & u_3 & f(u_1) & f(u_2) & f(u_3) \\ \hline \end{array}$$

$$\begin{matrix} \beta_1 = \\ \beta_2 = \\ \\ \end{matrix} \begin{bmatrix} 1 & 0 & -1 & 5 & -2 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & 0 & 3 & -2 & 1 \end{bmatrix} \qquad (7)$$

Comparing the right half of (6) with the left half of (7), we see that the space spanned by $f(u_1)$, $f(u_2)$ and $f(u_3)$ has as a basis

$$\beta_1 = u_1 - u_3$$

and (8)

$$\beta_2 = u_2 + 2u_3$$

We also have from (7) that

$$\left.\begin{array}{l} \beta_1 = 5f(u_1) - 2f(u_2) \\ \beta_2 = -2f(u_1) + u_2 \end{array}\right\}. \qquad (9)$$

Hence by the linearity of f we see from (9) that

$$\begin{array}{l} \beta_1 = f(5u_1 - 2u_2) \\ \beta_2 = f(-2u_1 + u_2) \end{array} \qquad (10)$$

In other words, to emphasize the domain of f, we see from (10) that if we define α_1 and $\alpha_2 \varepsilon V$ by

3.4.8(L) continued

$$\left.\begin{aligned} \alpha_1 &= 5u_1 - 2u_2 \\ \alpha_2 &= -2u_1 + u_2 \end{aligned}\right\} \tag{11}$$

then

$$f(v) = [f(\alpha_1),\ f(\alpha_2)]$$

<u>and</u>

$\{\alpha_1, \alpha_2\}$ is linearly independent.

<u>Important Aside</u>:
If $T:V \to W$ is any linear transformation of V into W and if $v_1, \ldots, v_n \varepsilon V$ are a set of linearly independent vectors we <u>cannot</u> be sure that $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent. In the present exercise, for example, $\{u_1, u_2, u_3\}$ is linearly independent but $\{f(u_1), f(u_2), f(u_3)\}$ isn't since dim $f(v) = 2$. On the other hand if $\{f(v_1), \ldots, f(v_n)\}$ is linearly independent so also is $\{v_1, \ldots, v_n\}$. Namely, suppose

$$c_1v_1 + \ldots + c_nv_n = 0 \tag{12}$$

then

$$f(c_1v_1 + \ldots + c_nv_n) = f(0) = 0,$$

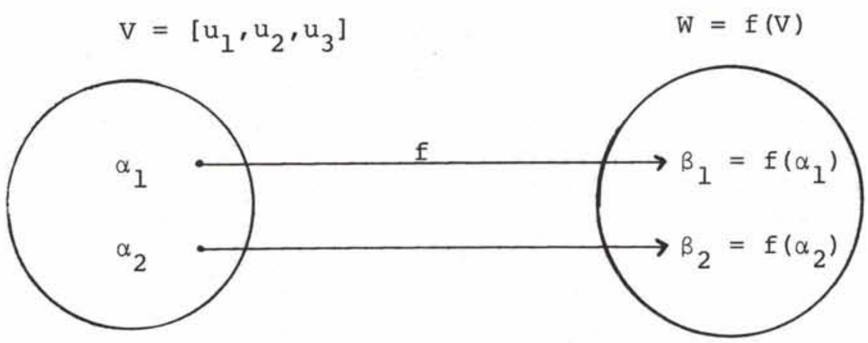
or

$$c_1f(v_1) + \ldots + c_nf(v_n) = 0. \tag{13}$$

Hence, since $\{f(v_1), \ldots, f(v_n)\}$ is linearly independent, we see from (13) that $c_1 = .. = c_n = 0$; and combining this with equation (12) we may conclude that $\{v_1, \ldots, v_n\}$ is linearly independent.

Summarized pictorially, we have

3.4.8(L) continued



$\alpha_1 = 5u_1 - 2u_2$
$\alpha_2 = -2u_1 + u_2$

1. $\{\alpha_1, \alpha_2\}$ is linearly independent
2. $f(v) = [f(\alpha_1), f(\alpha_2)]$

Suppose we let V_1 be the 2-dimensional subspace of V spanned by α_1 and α_2. Then, the only member of V_1 which belongs to the null space, N_f, of F is 0. Namely, if

$f(v) = 0$ and $v = c_1\alpha_1 + c_2\alpha_2$

then

$f(c_1\alpha_1 + c_2\alpha_2) = 0$

or

$c_1 f(\alpha_1) + c_2 f(\alpha_2) = 0$;

but since $\{f(\alpha_1), f(\alpha_2)\}$ is linearly independent, we have that $c_1 = c_2 = 0$, and hence, that $v = c_1\alpha_1 + c_2\alpha_2 = 0$.

In other words, if f_1 is the function f restricted to the domain V_1, $f_1 : V_1 \to W$ is both 1-1 and onto.*

*See **Note** at the end of this solution for more details.

3.4.8(L) continued

(Recall that in the lecture we showed that the linear transformation $T:V \to W$ was 1-1$\leftrightarrow$ $N_T = \{0\}$; i.e., $T(v) = 0 \leftrightarrow v = 0$.)

The fact that $V_1 \cap N_f = \{0\}$ now gives us a final hint to the structure of f(V). Namely, we know there must be an element $\alpha_3 \varepsilon V$ which does not belong to V_1 (since dim V = 3 and dim V_1 = 2). Moreover, α_3 must generate the null space of f since the zero element is the only element of the null space that belongs to V_1. To find a specific candidate for α_3 we need only look at the last row of (7) to conclude:

$$3f(u_1) - 2f(u_2) + f(u_3) = 0. \quad (14)$$

Hence, again by the linearity of f,(14) implies that

$$f(3u_1 - 2u_2 + u_3) = 0. \quad (15)$$

Looking at (15) we decide to let

$$\alpha_3 = 3u_1 - 2u_2 + u_3. \quad (16)$$

Clearly, $\alpha_3 \neq 0$ (since $\{u_1,u_2,u_3\}$ is linearly independent and $f(\alpha_3)\varepsilon N_f$ [since $f(\alpha_3) = 0$ by (15)]. Therefore, since

$N_f \cap V_1 = \{0\}$

and

$\alpha_3 \neq 0$,

we conclude that $\alpha_3 \notin V_1$. Hence,$\{\alpha_1,\alpha_2,\alpha_3\}$ is linearly independent. We also observe that the null space of V with respect to f consists of all scalar multiples of α_3. Namely, if $f(c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3) = 0$, then $c_1f(\alpha_1) + c_2f(\alpha_2) + c_3f(\alpha_3) = 0$, or since $f(\alpha_3) = 0$,

$$c_1f(\alpha_1) + c_2f(\alpha_2) = 0. \quad (17)$$

3.4.8(L) continued

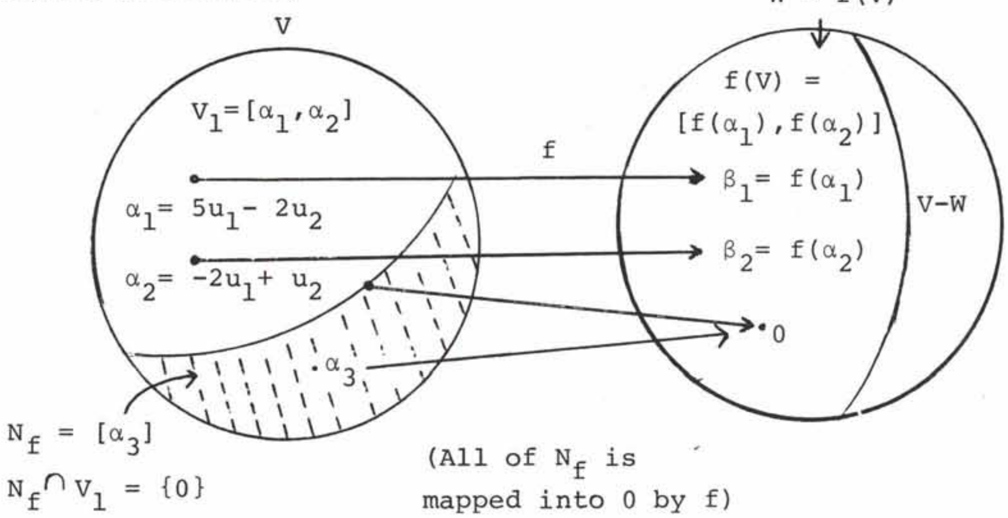
Since $\{f(\alpha_1), f(\alpha_2)\}$ is linearly independent, (17) implies that $c_1 = c_2 = 0$; therefore,

$$f(c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3) = 0 \leftrightarrow c_1 = c_2 = 0.$$

Hence,

$$N_f = \{c_3\alpha_3 : c_3 \varepsilon R\}$$

Again, pictorially



1. All multiples of α_3 map into 0.
2. The rest of f(V) comes from V_1.

Structurally what we are saying is that if f_1 is the restriction of f to V then $f(V) = f_1(V_1)$ and

$f_1 : V_1 \to f(V)$ is 1-1 and onto, but $f: V \to V$ is neither 1-1 nor onto.

Moreover, if $\alpha \varepsilon V_1$, then $f(\alpha) = f(\beta) \leftrightarrow (\beta - \alpha) \varepsilon N_f \leftrightarrow \beta = \alpha + \eta$ where $\eta \varepsilon N_f$. Namely, $f(\alpha) = f(\beta) \to f(\beta) - f(\alpha) = 0 \to f(\beta - \alpha) = 0$ $\beta - \alpha \varepsilon N_f$.

3.4.8(L) continued

In summary,

$$V = [\alpha_1, \alpha_2] \oplus [\alpha_3]$$

$$\dim [\alpha_1, \alpha_2] = \dim f(V) \ (= 2)$$

$$\dim [\alpha_3] = \dim N_f \ (= 1).$$

Hence, at least in this case,

$$\dim V = \dim f(V) + \dim N_f.$$

Optional Note:
We may outline the more general case as follows: Suppose $f:V \to W$ is linear and that

$$\dim V = n$$
$$\dim N_f = r \quad (\text{where } 0 \leq r \leq n)$$

Case 1:
$r = 0$, then $N_f = \{0\}$ so that f is 1-1 and onto. Consequently, V and W are "essentially" the same space (see the note on isomorphisms which follows).

Case 2:
$r = n$; then $f(V) = \{0\}$.

Case 3:
(The General Case; it includes cases 1 and 2 as special cases)

$$0 < r < n.$$

Let

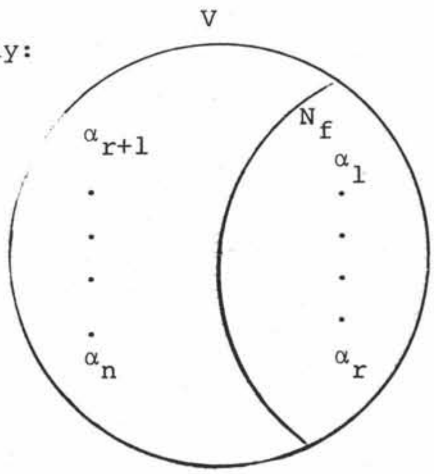
$$N_f = [\alpha_1, \ldots, \alpha_r].$$

Then since N_f is a subspace of V, we may find vectors $\alpha_{r+1}, \ldots,$ and $\alpha_n \varepsilon V$ but not in N_f such that

$$V = [\alpha_1, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_n].$$

3.4.8(L) continued

Pictorially:



Clearly $\{\alpha_1,\ldots,\alpha_n\}$ spans $f(V)$ for if $v \varepsilon V$, then

$$v = \sum_{i=1}^{n} c_i\alpha_i \; .$$

Hence,

$$f(v) = f\left(\sum_{i=1}^{n} c_i\alpha_i\right)$$

or by linearity

$$f(v) = \sum_{i=1}^{n} c_i f(\alpha_i) .$$

That is,

$$v = c_1\alpha_1 + \ldots + c_n\alpha_n \rightarrow$$
$$f(v) = c_1 f(\alpha_1) + \ldots + c_n f(\alpha_n) . \qquad (18)$$

Moreover, since $f(\alpha_1) = \ldots = f(\alpha_r) = 0$, we see from (18) that

$$f(v) = c_{r+1} f(\alpha_{r+1}) + \ldots + c_n f(\alpha_n) . \qquad (19)$$

Hence, from (19) we conclude that

$$\{f(\alpha_{r+1}),\ldots,f(\alpha_n)\}$$

span $f(V)$.

3.4.8(L) continued

All that remains to be shown is that $\{f(\alpha_{r+1}),\ldots,f(\alpha_n)\}$ is linearly independent, for if this is so,

$$f(V) = \overbrace{[f(\ _{r+1}),\ldots,f(\ _{n})]}^{n\ -\ r\ \text{elements}}$$

in which case

$$\dim V = \dim N_f + \dim f(V).$$

I.e.,

$$n = r + (n - r).$$

So assume

$$x_{r+1}f(\alpha_{r+1}) + \ldots + x_n f(\alpha_n) = 0. \qquad (20)$$

By linearity (20) implies

$$f(x_{r+1}\alpha_{r+1} + \ldots + x_n\alpha_n) = 0$$

so that

$$x_{r+1}\alpha_{r+1} + \ldots + x_n\alpha_n \varepsilon N_f.$$

Since $N_f = [\alpha_1,\ldots,\alpha_r]$, there exist constant $-x_1,\ldots,-x_r$* such that

$$x_{r+1}\alpha_{r+1} + \ldots + x_n\alpha_n = -x_1\alpha_1 + \ldots + -x_r\alpha_r.$$

Hence,

$$x_1\alpha_1 + \ldots + x_r\alpha_r + x_{r+1}\alpha_{r+1} + \ldots + x_n\alpha_n = 0. \qquad (21)$$

But since $\{\alpha_1,\ldots,\alpha_n\}$ is linearly independent, (21) implies that

*The reason for using the minus signs is for arithmetic convenience.

3.4.8(L) continued

$x_1 = \dots = x_n = 0$. In particular, then, $x_{r+1} = \dots = x_n$, so from (20) we see that $\{f(\alpha_{r+1}),\dots,f(\alpha_n)\}$ is linearly independent.

Note:
From (7) we have that

$$w = x_1u_1 + x_2u_2 + x_3u_3 \varepsilon f(V) \qquad (22)$$

$$w = x_1\beta_1 + x_2\beta_2$$

$$= x_1(u_1 - u_3) + x_2(u_2 + 2u_3)$$

$$= x_1u_1 + x_2u_2 + (-x_1 + 2x_2)u_3. \qquad (23)$$

Comparing (22) and (23), we see that

$$(x_1,x_2,x_3)\varepsilon f(V) \leftrightarrow x_3 = -x_1 + 2x_2.$$

Geometrically, then, if we view this problem as one in which xyz-space is mapped into uvw-space, we see that the image of our mapping is the plane

$w = -u + 2v$.

The scalar multiple of $3\vec{i} - 2\vec{j} + \vec{k}$ map into 0. That is, our null space is the line

$$\left.\begin{aligned} x &= 3t \\ y &= -2t \\ z &= t \end{aligned}\right\}$$

A NOTE ON ISOMORPHISM (OPTIONAL)

By this stage of the game, we should feel very much at home with the notion of a mathematical structure. In terms of our axiomatic approach, we have seen that the only valid conclusions are those which follow inescapably from our axioms. Hence, in this respect, we cannot mathematically (logically) distinguish between two different physical models if the only listed axioms are the same for both models. (In somewhat different perspective, any difference between the two models depends on additional axioms which have not been stated.)

As a very simplistic non-mathematical illustration, consider the plight of a champion chess player who is called upon to play a game in which the pieces have been disguised so that he does not recognize which piece is the king and which is the queen, etc. Notice that he cannot play the game now! For example, he can't even be sure as to how each piece is to be placed on the board. However, his difficulty is only temporary (hopefully), since once he learns the identity of the pieces, the game is the same as the one he is champion in. In other words, he soon discovers that except for how the pieces are named(in this case, how they look) the structure is the same as in the "regular" game. In short, the strategy of a game hinges on the relationship between the various terms (rules or axioms) rather than on the names of the terms themselves. An even easier example might be to observe that if a baseball buff did not know a single word of Japanese he could still enjoy a baseball game played by Japanese, in Japanese, in Japan.

Applying this idea to mathematics, we often agree to identify two systems as being the same if the only way we can tell them apart is by the name of the terms. For instance, relative to our present discussion of linear transformations of one vector space into another, we have already seen that such a mapping preserves structure. That is, a linear combination of elements in the domain of f is mapped into the same linear combination of the images of the elements in the domain. In other words, the image of

$$c_1v_1 + c_2v_2 + \dots + c_nv_n$$

is

$$c_1f(v_1) + c_2f(v_2) + \dots + c_nf(v_n).$$

Of course, f need not be either 1-1 or onto.

We know, however, that if the null space of f consists of 0 alone, then f is both 1-1 and onto. In this case we call f an isomorphism. This ominous term is just what the name implies. It implies that V and f(V) have the same form or structure.

Namely, any valid result involving a linear combination of $v_1, \ldots,$ and $v_n \varepsilon V$ remains valid in f(V) when $v_1, \ldots, v_n$ are replaced by $f(v_1), \ldots, f(v_n)$.

As an example, consider the set of integers J relative to the structure of addition, and consider the function which doubles a given integer. In other words, if f(x) = 2x for each $x \varepsilon J$ and if E is the subset of J consisting of all even integers, then

$$f: J \to E$$

is both 1-1 and onto.

But much more than this, J and E are isomorphic as far as addition is concerned since

$$\begin{aligned} f(x+y) &= 2(x+y) \\ &= 2x + 2y \\ &= f(x) + f(y). \end{aligned}$$

What this means is that if we replace integer x by its double 2x, any additional fact which is valid in J remains valid in E.

Diagramatically, if we line up elements of J with the corresponding element of E:

-2	-1	0	1	2	+	3	4	5	6	7	8
-4	-2	0	1	4	+	6	8	10	12	14	16

That is $2 + 3 = 5 \leftrightarrow f(2) + f(3) = f(5)$, etc.

Notice that it is not enough for f to be 1-1 and onto; we also require that f preserve structure. For example, with J, E, and f as above notice that f does not preserve the structure of J relative to multiplication. Namely,

$f(xy) = 2xy$, while $f(x)f(y) = (2x)(2y) = 4xy$. Hence $f(xy \neq f(x)f(y)$ [i.e., $f(x)f(y) = 2f(xy)$].

Again, diagramatically

These do not "match".

In other words, we would say that the integers and even integers are isomorphic with respect to the structure of addition but they are not isomorphic with respect to the structure of multiplication.

A complete discussion on isomorphism is beyond the aim of the present course, but as a closing aside, we can now explain in terms of isomorphism why one usually writes n-dimensional vector spaces as n-tuples and often fails to talk about "coordinate systems". Namely, if dim $V = n$ and if we specify a particular basis for V, say, $V = [v_1,\ldots,v_n]$; then <u>relative to the particularly chosen basis</u>, V is isomorphic to E^n. More specifically (and we have done this several times informally in this Block) given v in V we know that v can be written uniquely in the form

$$v = c_1v_1 + \ldots + c_nv_n \tag{1}$$

and we define $f:V \to E^n$ by

$$f(v) = (c_1,\ldots,c_n) \tag{2}$$

where $c_1,\ldots,c_n$, and v are as in (1). It is then trivially established that f is 1-1 and onto and also that f is linear. In our informal approach we said all of this simply by saying "Let $(c_1,\ldots,c_n)$ be an <u>abbreviation</u> for $c_1v_1 + \ldots + c_nv_n$". Certainly, the connotation of "abbreviation" seems to be that it is a different (shorter) way of saying the <u>same</u> thing.

What we must be careful about, however, (as we have also emphasized in other lectures) is that two different bases lead to different structures of E^n. That is, $(c_1, \ldots, c_n)$ names one element with respect to one basis and (possibly) a different element with respect to another basis. If V and W are isomorphic, we write $V \cong W$.

3.4.9 (optional)

a. Here we have generalized the previous exercise in the sense that $W \neq V$. To find $f(V)$ [and notice that this is a subspace of W not V; in the previous exercise, this distinction wasn't so apparent since then $V = W$] we want to find the space spanned by $f(v_1)$, $f(v_2)$, $f(v_3)$, and $f(v_4)$. This leads to

$$\begin{array}{cc|cccc} w_1 & w_2 & f(v_1) & f(v_2) & f(v_3) & f(v_4) \\ \hline \end{array}$$

$$\left[\begin{array}{cc|cccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 & 0 \\ 3 & 5 & 0 & 0 & 1 & 0 \\ 4 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \sim$$

$$\left[\begin{array}{cccccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 2 & -3 & 0 & 1 & 0 \\ 0 & -3 & -4 & 0 & 0 & 1 \end{array}\right] \sim$$

$$\left[\begin{array}{cc|cccc} 1 & 0 & 3 & -1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & -10 & 3 & 0 & 1 \end{array}\right] \quad (1)$$

From (1) we deduce that

$$w_1 = 3f(v_1) - f(v_2) = f(3v_1 - v_2) \quad (2)$$

$$w_2 = -2f(v_1) + f(v_2) = f(-2v_1 + v_2) \quad (3)$$

$$0 = f(v_1) - 2f(v_2) + f(v_3) = f(v_1 - 2v_2 + v_3) \quad (4)$$

$$0 = -10f(v_1) + 3f(v_2) + f(v_4) = f(-10v_1 + 3v_2 + v_4). \quad (5)$$

3.4.9 continued

From (2) and (3) we see that

$$f(V) = [f(\alpha_1), f(\alpha_2)] = W \tag{6}$$

$$\alpha_1 = 3v_1 - v_2 \tag{7}$$

$$\alpha_2 = -2v_1 + v_2. \tag{8}$$

That is, $f:V \to W$ is onto; in particular $\dim f(V) = 2$. From (4) and (5) we deduce that

$$N_f = [\alpha_3, \alpha_4], \tag{9}$$

where

$$\alpha_3 = v_1 - 2v_2 + v_3 \tag{10}$$

$$\alpha_4 = -10v_1 + 3v_2 + v_4. \tag{11}$$

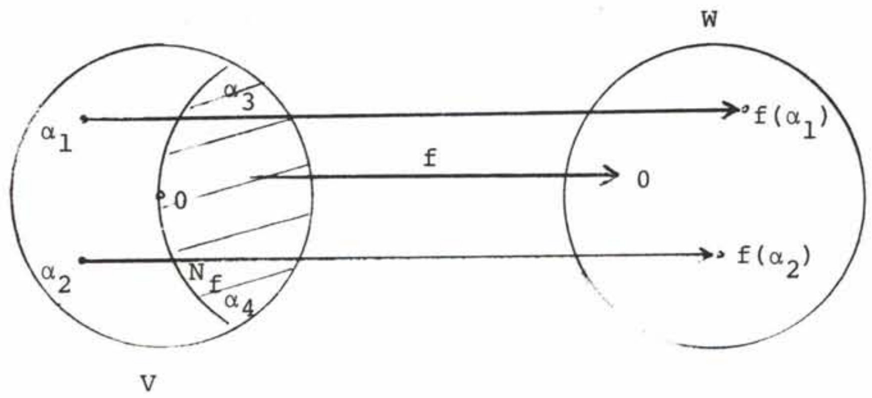
Thus, if we now let

$V_1 = [\alpha_1, \alpha_2]$ and $N_f = [\alpha_3, \alpha_4]$,

we have

1. $V = V_1 + N_f$
2. If f_1 is the restriction of f to V_1, then $f_1:V_1 \cong W$.

That is, $f_1:V_1 \to W$ is 1-1 and onto. Pictorially,



3.4.9 continued

b. To obtain a basis for N_f in row-reduced form we have

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ -10 & 3 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & -17 & 10 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 17 & -34 & 17 & 0 \\ 0 & 34 & -20 & -2 \end{bmatrix}$$

$$\sim \begin{bmatrix} 17 & 0 & -3 & -2 \\ 0 & 34 & -20 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -\frac{3}{17} & -\frac{2}{17} \\ 0 & 1 & -\frac{10}{17} & -\frac{1}{17} \end{bmatrix} \qquad (12)$$

From (12),

$N_f = [\beta_1, \beta_2]$

where

$$\beta_1 = v_1 - \frac{3}{17} v_3 - \frac{2}{17} v_4$$
$$\beta_2 = v_2 - \frac{10}{17} v_3 - \frac{1}{17} v_4 \qquad (13)$$

Hence,

$N_f = \{ x_1\beta_1 + x_2\beta_2 : x_1, x_2 \varepsilon R \}$,

so by (13),

$$N_f = \{(x_1v_1 - \frac{3}{17} x_1v_3 - \frac{2}{17} x_1v_4) + (x_2v_2 - \frac{10}{17} x_2v_3 - \frac{1}{17} x_2v_4)\}$$

$$= \{ x_1v_1 + x_2v_2 + [-\frac{3}{17} x_1 - \frac{10}{17} x_2]v_3 + [-\frac{2}{17} x_1 - \frac{1}{17} x_2]v_4\}$$

3.4.9 continued

Hence, relative to v_1, v_2, v_3, v_4 as a coordinate system

$$(x_1, x_2, x_3, x_4) \in N_f \leftrightarrow \begin{cases} x_3 = -\frac{3}{17} x_1 - \frac{10}{17} x_2 \\ x_4 = -\frac{2}{17} x_1 - \frac{1}{17} x_2 \end{cases} .$$

In other words,

$$x_1v_1 + x_2v_2 + x_3v_3 + x_4v_4 \in N_f \leftrightarrow$$

$$\left.\begin{aligned} 3x_1 + 10x_2 + 17x_3 &= 0 \\ 2x_1 + x_2 + 17x_4 &= 0 \end{aligned}\right\}$$

<u>Check</u>:

$\alpha_3 = v_1 - 2v_2 + v_3$.

Hence, $x_1 = 1$, $x_2 = -2$, $x_3 = 1$, $x_4 = 0$, whereupon

$$3x_1 + 10x_2 + 17x_3 = 3 - 20 + 17 = 0$$
$$2x_1 + x_2 + 17x_4 = 2 - 2 + 0 = 0$$

$\alpha_4 = -10v_1 + 3v_2 + v_4$.

Hence, $x_1 = -10$, $x_2 = 3$, $x_3 = 0$, $x_4 = 1$, whereupon

$$3x_1 + 10x_2 + 17x_3 = -30 + 30 + 0 = 0$$
$$2x_1 + x_2 + 17x_4 = -20 + 3 + 17 = 0.$$

In other words, relative to $\{v_1, v_2, v_3, v_4\}$ as our coordinate system,

$(1,0, -\frac{3}{17}, -\frac{2}{17})$ and $(0,1,-\frac{10}{17}, -\frac{1}{17})$

form a basis for N_f so that any element of N_f has the form

3.4.9 continued

$$x(1,0,-\frac{3}{17},-\frac{2}{17})+y(0,1,-\frac{10}{17},-\frac{1}{17})$$

$$=(x,y,-\frac{3x+10y}{17},-\frac{2x+y}{17}). \tag{14}$$

c. $f(x_1v_1+x_2v_2+x_3v_3+x_4v_4)=5w_1+6w_2.$

Hence, by (2) and (3)

$$f(x_1v_1+x_2v_2+x_3v_3+x_4v_4)=5f(3v_1-v_2)+6f(-2v_1+v_2) \tag{15}$$

$$=f(15v_1-5v_2)+f(-12v_1+6v_2) \tag{16}$$

$$=f(3v_1+v_2) \tag{17}$$

(where (16) and (17) follow from (15) by the linearity of f).

Hence, one $v\varepsilon V$ such that $f(v)=5w_1+6w_2$ is $v=3v_1+v_2$.

<u>Check</u>:

$$\begin{aligned}f(3v_1+v_2)&=3f(v_1)+f(v_2)\\&=3(w_1+w_2)+(2w_1+3w_2)\\&=5w_1+6w_2.\end{aligned}$$

In n-tuple notation, using $\{v_1,v_2,v_3,v_4\}$ as a coordinate system, (3,1,0,0) maps into $5w_1+6w_2$.

Now since $f(v)=0$ for every $v\varepsilon N_f$, we conclude from (14) that

$$f(3,1,0,0)+f(x,y,-\frac{3x+10y}{17},-\frac{2x+y}{17})=0$$

or

$$f(3+x,1+y,-\frac{3x-10y}{17},-\frac{2x-y}{17})=0.$$

In other words, the set of all $v\varepsilon V$ such that

$$v=(3+x,1+y,-\frac{3x-10y}{17},-\frac{2x-y}{17})$$

has the property that $f(v)=5w_1+6w_2$.

3.4.9 continued

In summary,

$\gamma = 3v_1 + v_2$

is the only member of $V_1 = [\alpha_1,\alpha_2] = [v_1,v_2]$ such that

$f(\gamma) = 5w_1 + 6w_2$.

The set of all $v \varepsilon V$ such that $f(v) = 5w_1 + 6w_2$ is then given by

$\{\gamma + \eta : \eta \varepsilon N_f\}$.

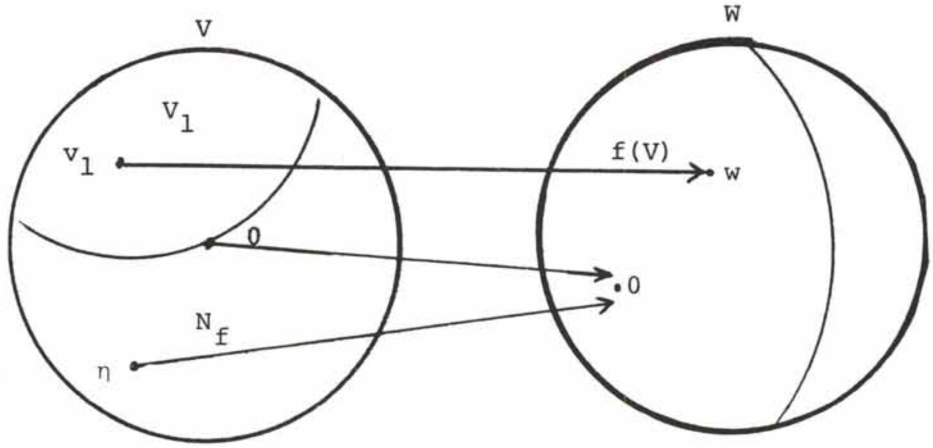
More generally, if $N_f = [\alpha_1,\ldots,\alpha_r]$ and $V = [\alpha_1,\ldots,\alpha_r,\alpha_{r+1},\ldots,\alpha_n]$, define V_1 by $[\alpha_{r+1},\ldots,\alpha_n]$. Then, if f_1 is f restricted to V_1,

$f_1 : V_1 \cong f(V)$.

That is, for $w \varepsilon f(V)$ there exists one and only one $u \varepsilon V_1$ such that $f_1(u) = f(u) = w$. We then find all $v \varepsilon V$ such that $f(v) = w$ by adding any element of N_f to u. That is

$\{v : f(v) = w\} = \{u + \eta : \eta \varepsilon N_f\}$.

Pictorially



3.4.9 continued

$f(v_1) = w$
$\underline{f(\eta) = 0}$
$f(v_1) + f(\eta) = w,$
therefore $f(v_1 + \eta) = w$.

1. For $w \epsilon f(V)$ there exists a unique $v_1 \epsilon V_1$ such that $f(v_1) = w$.
2. The set of all $v \epsilon V$ such that $f(v) = w$ is then given by $\{v_1 + \eta : \eta \epsilon N_f\}$.

3.4.10 (optional)

Suppose that $f: V \to V$ is a linear transformation of V into itself and that W is a subspace of V. Then we already know that f(W) must be a subspace of f(V) = V. What need not be true, however, is that f(W) must be a subspace of W. For example, consider the linear mapping of the plane which maps $\vec{i}$ into $\vec{i} + \vec{j}$ and $\vec{j}$ into $\vec{i} - \vec{j}$. This mapping carries the 1-dimensional subspace (i.e., the line) y = 0 onto the 1-dimensional subspace y = x, but certainly the lines y = 0 and y = x are different subspaces of the plane.

If it happens that f(W) is a subspace of W then we refer to W as being an <u>invariant subspace</u> of V with respect to f. Without going into any detail, it should be clear that one is often happy to deal with invariant subspaces. That is, it's nice to know what subspaces are <u>preserved</u> by the given linear transformation.

The aim of this exercise is to show that for **any** given constant c, the solutions of the equations

$$f(v) = cv \qquad (1)$$

are not only a subspace of V but are an invariant subspace relative to f.

To this end, suppose

$$w = \{v \epsilon V : f(v) = cv\} \qquad (2)$$

where in (1), c is a <u>fixed</u> constant.

3.4.10 continued

Then for v_1 and $v_2 \varepsilon W$ we have

$$\left.\begin{array}{l} f(v_1) = cv_1 \\ f(v_2) = cv_2 \end{array}\right\} . \qquad (3)$$

Hence, by the linearity of f, we see from (2) that

$$\begin{aligned} f(v_1 - v_2) &= cv_1 - cv_2 \\ &= c(v_1 - v_2). \end{aligned}$$

Therefore, by definition of W,

$$v_1 - v_2 \varepsilon W. \qquad (4)$$

Moreover, for any scalar k and any vector $w \varepsilon W$ we have

$$\begin{aligned} f(kw) &= kf(w) \\ &= k(cw) \\ &= c(kw), \end{aligned}$$

so again by the definition of W,

$$kw \varepsilon W. \qquad (5)$$

From (3) and (4) we conclude that

$$\left.\begin{array}{l} v_1,\ v_2 \varepsilon W \rightarrow v_1 - v_2 \varepsilon W \\ \text{and} \\ k \varepsilon R,\ w \varepsilon W \rightarrow kw \varepsilon W \end{array}\right\} .$$

Consequently W is a subspace of V.

To prove that W is an invariant subspace of V relative to f, we must show that $w \varepsilon W \rightarrow f(w) = W$. So suppose $w \varepsilon W$. Then,

$$f(w) = cw. \qquad (6)$$

3.4.10 continued

But since W is itself a vector space, $w \in W$ implies that $cw \in W$. Hence from (5) we see that

$$w \in W \to f(w)\ [= cw] \in W.$$

Hence W is invariant with respect to f.

Note #1:

Observe that the notion of invariant subspace depends on the particular linear transformation being considered. For example if W is a subspace of V, certainly this fact does not depend on the transformation f. Namely, the study of subspaces of a given space is independent of the notion of mappings. However, it is equally clear that what the image of W is does depend on the mapping being considered. Relative to our earlier remarks the lines $y = 0$ and $y = x$ are both subspaces of the plane, but what their images are with respect to a linear mapping depends on the particular mapping. In summary, a subspace W of V may be invariant to one linear mapping $f:V \to V$ but not invariant with respect to another linear mapping $g:V \to V$.

Note #2:

From a geometric point of view, the vectors defined by (1) are those which have their direction preserved with respect to a given linear transformation. If a basis consisting of such vectors exists, then this is a very nice basis to use for this transformation since then the basis vectors of W are also basis vectors for f(W). In other words, relative to this basis, our coordinate axes are preserved under the transformation. We shall illustrate this in more detail in Exercise 3.6.3(L) in our discussion of Eigenvectors.