chapter II Linear transformations and matrices

In this chapter we define linear transformations and various operations: addition of two linear transformations, multiplication of two linear transformations, and multiplication of a linear transformation by a scalar. Linear transformations are functions of vectors in one vector space U with values which are vectors in the same or another vector space V which preserve linear combinations. They can be represented by matrices in the same sense that vectors can be represented by n-tuples. This representation requires that operations of addition, multiplication, and scalar multiplication of matrices be defined to correspond to these operations with linear transformations. Thus we establish an algebra of matrices by means of the conceptually simpler algebra of linear transformations.

The matrix representing a linear transformation of U into V depends on the choice of a basis in U and a basis in V. Our first problem, a recurrent problem whenever matrices are used to represent anything, is to see how a change in the choice of bases determines a corresponding change in the matrix representing the linear transformation. Two matrices which represent the same linear transformation with respect to different sets of bases must have some properties in common. This leads to the idea of equivalence relations among matrices. The exact nature of this equivalence relation depends on the bases which are permitted.

In this chapter no restriction is placed on the bases which are permitted and we obtain the widest kind of equivalence. In Chapter Ill we identify U and V and require that the same basis be used in both. This yields a more restricted kind of equivalence, and a study of this equivalence is both interesting and fruitful. In Chapter V we make further restrictions in the permissible bases and obtain an even more restricted equivalence.

When no restriction is placed on the bases which are permitted, the

26

equivalence is so broad that it is relatively uninteresting. Very useful results are obtained, however, when we are permitted to change basis only in the image space V. In every set of mutually equivalent matrices we select one, representative of all of them, which we call a normal form, in this case the Hermite normal form. The Hermite normal form is one of our most important and effective computational tools, far exceeding in utility its application to the study of this particular equivalence relation.

The pattern we have described is worth conscious notice since it is recurrent and the principal underlying theme in this exposition of matrix theory. We define a concept, find a representation suitable for effective computation, change bases to see how this change affects the representation, and then seek a normal form in each class of equivalent representations.

1 1 Linear Transformations

Let U and V be vector spaces over the same field of scalars F.

Definition. A linear transformation of U into V is a single-valued mapping of U into V which associates to each element e U a unique element c(oc) e V such that for all cc, G U and all a, b e F we have

 + bß) = ac(oc) + bc(ß). (1.1)

We call c(æ) the image of under the linear transformation c. If e V, then any vector e U such that c(oc) = is called an inverse image of a. The set of all oc e U such that = is called the complete inverse image of a, and it is denoted by Generally, need not be a single element as there may be more than one e U such that c(u) = a.

By taking particular choices for a and b we see that for a linear transformation + F) = c(oc) + c(ß) and o(aoc) = ac(oc). Loosely speaking, the image of the sum is the sum of the images and the image of the product is the product of the images. This descriptive language has to be interpreted generously since the operations before and after applying the linear transformation may take place in different vector spaces. Furthermore, the remark about scalar multiplication is inexact since we do not apply the linear transformation to scalars; the linear transformation is defined only for vectors in U. Even so, the linear transformation does preserve the structural operations in a vector space and this is the reason for its importance. Generally, in algebra a structure-preserving mapping is called a homomorphism. To describe the special role of the elements of F in the condition, c(aoc) — ac(u), we say that a linear transformation is a homomorphism over F, or an F-homomorphism.

If for it necessarily follows that o(oc) c(ß), the homomorphism c is said to be one-to-one and it is called a monomorphism. If A is any subset of U, c(A) will denote the set of all images of elements of A; c(A) = { a l ä = o(u) for some e A}. c(A) is called the image of A. c(U) is often denoted by Im(c) and is called the image of c. IfIm(5) = V we shall say that the homomorphism is a mapping onto V and it is called an epimorphism.

We call the set U, on which the linear transformation c is defined, the domain of o. We call V, the set in which the images of c are defined, the codomain of c. Strictly speaking, a linear transformation must specify the domain and codomain as well as the mapping. For example, consider the linear transformation that maps every vector of U onto the zero vector of V. This mapping is called the zero mapping. If W is any subspace of V, there is also a zero mapping of U into W, and this mapping has the same effect on the elements of U as the zero mapping of U into V. However, they are different linear transformations since they have different codomains. This may seem like an unnecessarily fine distinction. Actually, for most of this book we could get along without this degree of precision. But the more deeply we go into linear algebra the more such precision is needed. In this book we need this much care when we discuss dual spaces and dual transformations in Chapter IV.

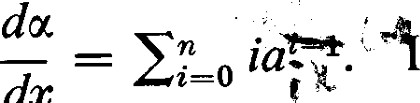
A homomorphism that is both an epimorphism and a monomorphism is called an isomorphism. If e V, the fact that c is an epimorphism says that —here is an e U such that c(oc) = ä. The fact that c is a monomorphism says that this is unique. Thus, for an isomorphism, we can define an inverse mapping that maps onto u.

Theorem 1.1. The inverse of an isomorphism is also an isomorphism. PROOF. Since is obviously one-to-one and onto, it is necessary only to show that it is linear. If = c(ot) and Ü(ß), then a(aæ + b") aä + b/ so that + bB) = au + bß = ac-l (ä) + bo I (ß). [Z

For the inverse isomorphism is an element of U. This conflicts with the previously given definition of c—I (ä) as a complete inverse image in which o-l (ä) is a subset of U. However, the symbol c-1 , standing alone, will always be used to denote an isomorphism, and in this case there is no diffculty caused by the fact that c-l (ä) might denote either an element or a oneelement set.

Let us give some examples of linear transformations. Let (J = V = P the space of polynomials in x with coeffcients in R. For

define 0(1) — n calculus one of the very first things + P) doc dÅ proved about the derivative is that it is linear, dc dc + — dc and



dx

d(aæ)  The mapping T(u) = o cc+1 is also linear.  Notice dc that this is not the indefinite integral since we have specified that the constant of integration shall be zero. Notice that c is onto but not one-to-one and T is one-to-one but not onto.

Let U = RO and V = R m with m n. For each oc = (al, . . . , n define c(oc) = (al, . . . , am) e Rm . It is clear that this linear transformation is one-to-one if and only if m = n, but it is onto. For each = (bl, bm) e R rn define T(ß)  O) e P.n . This linear transformation is one-to-one, but it is onto if and only if m = n.

Let U = V. For a given scalar a e F the mapping of oc onto au is linear since



and a(boc) = (ab)oc = (ba)oc = b • a(oc).

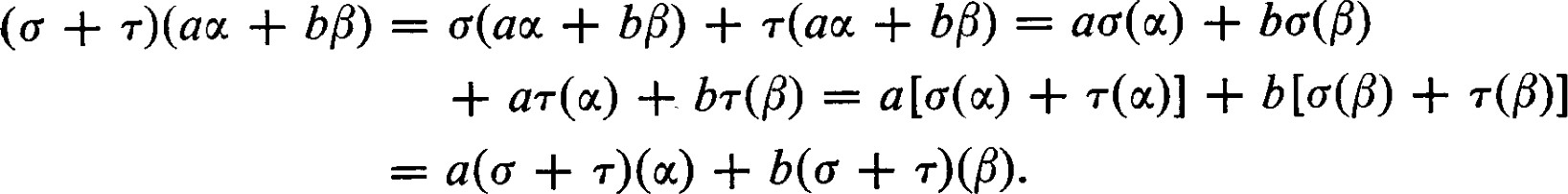
To simplify notation we also denote this linear-transformation by a. Linear transformations of this type are called scalar transformations, and there is a one-to-one correspondence between the field of scalars and the set of scalar transformations. In particular, the linear transformation that leaves every vector fixed is denoted by l. It is called the identity transformation or unit transformation. If linear transformations in several vector spaces are being discussed at the same time, it may be desirable to identify the space on which the identity transformation is defined. Thus IU will denote the identity transformation on U.

When a basis of a finite dimensional vector space V is used to establish a correspondence between vectors in V and n-tuples in Fn , this correspondence is an isomorphism. The required arguments have already been given in Section 1-3. Since V and Fn are isomorphic, it is theoretically possible to discuss the properties of V by examining the properties of Fn . However, there is much interest and importance attached to concepts that are independent of the choice of a basis. If a homomorphism or isomorphism can be defined uniquely by intrinsic properties independent of a choice of basis the mapping is said to be natural or canonical. In particular, any two vector spaces of dimension n over F are isomorphic. Such an isomorphism can be established by setting up an isomorphism between each one and Fn . This isomorphism will be dependent on a choice of a basis in each space. Such an isomorphism, dependent upon the arbitrary choice of bases, is not canonical.

Next, let us define the various operations between linear transformations. For each pair c, T of linear transformation of U into V, define + T by the rule



c + T is a linear transformation since



# Observe that addition of linear transformation is commutative; c + T

For each linear transformation c and a e F define ac by the rule; a [c(oc)]. ac is a linear transformation.

It is not diffcult to show that with these two operations the set of all linear transformations of U into V is itself a vector space over F. This is a very important fact and we occasionally refer to it and make use of it. However, we wish to emphasize that we define the sum of two linear transformations if and only if they both have the same domain and the same codomain. It is neither necessary nor suffcient that they have the same image, or that the image of one be a subset of the image of the other. It is simply a question of being clear about the terminology and its meaning. The set of all linear transformations of U into V will be denoted by Hom(U, V).

There is another, entirely new, operation that we need to define. Let W be a third vector space over F. Let c be a linear transformation of U into V and T a linear transformation of V into W. By TO we denote the linear transformation of U into W defined by the rule: = Notice that in this context has no meaning. We refer to this operation as either iteration or multiplication of linear transformation, and is called the product of T and c.

The operations between linear transformations are related by the following rules :

l. Multiplication is associative: afro) = (TT)c. Here is a linear transformation of W into a fourth vector space X.

1. Multiplication is distributive with respect to addition:

(Tl -k T2)ff = TIC + and T(CI + 62) = TCI + TC2.

1. Scalar multiplication commutes with multiplication: afro) = T(ac). These properties are easily proved and are left to the reader.

Notice that if W # U, then is defined but CT is not. If all linear transformations under consideration are mappings of a vector space U into itself, then these linear transformations can be multiplied in any order. This means that and would both be defined, but it would not mean that = CT.

The set of linear transformation of a vector space into itself is a vector space, as we have already observed, and now we have defined a product which satisfies the three conditions given above. Such a space is called an associative algebra. In our case the algebra consists of linear transformation and it is known as a linear algebra. However, the use of terms is always in a state of flux, and today this term is used in a more inclusive sense. When referring to a particular set with an algebraic structure, "linear algebra" still denotes what we have just described. But when referring to an area of study, the term "linear algebra includes virtually every concept in which linear transformations play a role, including linear transformations between different vector spaces (in which the linear transformations cannot always be multiplied), sequences of vector spaces, and even mappings of sets of linear transformations (since they also have the structure of a vector space).

Theorem 1.2. Im(c) is a subspace of V.

PROOF. If and B are elements of Im(c), there exist I, e U such that c(oc) = and c(ß) = B. For any a, b e F, c(aoc + bÅ) = ac(æ) + bc(ß) — aä + b/ e Im(c). Thus Im(c) is a subspace of V. D

Corollary 1.3. If UI is a subspace of U, then 6(U1) is a subspace of V. 

It follows from this corollary that c(O) = 0 where 0 denotes the zero vector of U and the zero vector of V. It is even easier, however, to show it directly. Since 6(0) = + 0) = 0(0) + 0(0) it follows from the uniqueness of the zero vector that c(0) = O.

For the rest of this book, unless specific comment is made, we assume that all vector spaces under consideration are finite dimensional. Let dim U = n and dim V = m.

The dimension of the subspace Im(c) is called the rank of the linear transformation c. The rank of c is denoted by p(c).

Theorem 1.4. {m, n}.

PROOF. If {WI, us} is linearly dependent in U, there exists a non-trivial relation of the form i apti = O. But then ai0(q) = c(0) = O; that is,  , c(æs)} is linearly dependent in V. A linear transformation preserves linear relations and transforms dependent sets into dependent



sets. Thus, there can be no more than n linearly independent elements in Im(c). In addition, Im(c) is a subspace of V so that dim Im(c) m. Thus p(c) = dim Im(o) min {m, n). 

Theorem 1.5. If W is a subspace of V, the set rI (W) of all e U such that c(oc) e W is a subspace of U.

PROOF. If u, e c-I (W), then c(aoc + bß) = ac(u) + bc(ß) e W. Thus au + bf e and is a subspace. 

The subspace K(c) = is called the kernel of the linear transformation c. The dimension ofK(c) is called the nullity Ofc. The nullity of c is denoted by v(c).

Theorem 1.6. p(c) + v(«)

PROOF. Let WI, . , 1,} be a basis of U such that {0%, is a basis of For oc — apti + bißj e IJ we see that c(u) =

+ bjc(ßj) = Thus {c(ßl) c(h)} spans Im(c). On the other hand if = O, then c(Ei cßj) = cjc(ßj) = O; that is, 2,. cjßj e In this case there exist coefficients dt such that cjß,  di0(i. If any of these coeffcients were non-zero we would have a nontrivial relation among the elements of {11, 01 13 }. Hence, all  = 0 and {c(ßl), c(ßk)} is linearly independent. But then it is a basis of Im(o) so that k = p(c). Thus p(c) + v(c) = n. 

Theorem 1.6 has an important geometric interpretation. Suppose that a 3-dimensional vector space R3 were mapped onto a 2-dimensional vector space R2 . In this case, it is simplest and suffciently accurate to think of c as the linear transformation which maps (al, '12, '13) e R3 onto (al, ao e R2 which we can identify with (al, a2, 0) e R3 . Since p(c)

Clearly, every point (0, 0, ao on the 4-axis is mapped onto the origin. Thus K(Ü) is the A-axis, the line through the origin in the direction of the projection, and {(0, 0, l) = 11} is a basis of It should be evident that any plane through the origin not containing K(Ü) will be projected onto the CIC2-plane and that this mapping is one-to-one and onto. Thus the complementary subspace (61, 62) can be taken to be any plane through the origin not containing the 4-axis. This illustrates the wide latitude of choice possible for the complementary subspace WI,

Theorem 1.7. A linear transformation c of U into V is a monomorphism if and only ifv(c) = 0, and it is an epimorphism if and only if p(c) = dim V.

PROOF. K(Ü) = {0} if and only if v(c) = O. If 6 is a monomorphism, then certainly K(c) = {()} and v(c) = 0. On the other hand, if v(Ü) = 0 and c(u) = c(ß), then c(oc — 16) = 0 so that — = {0}. Thus, if v(o) = 0, is a monomorphism.

It is but a matter of reading the definitions to see that c is an epimorphism if and only if p(o) = dim V. 

If dim U = n < dim V = m, then p(c) = n — v(c) n < m so that c cannot be an epimorphism. Ifn > m, then v(c) = n — p(c) n — m > 0, so that c cannot be a monomorphism. Any linear transformation from a vector space into a vector space of higher dimension must fail to be an epimorphism. Any linear transformation from a vector space into a vector space of lower dimension must fail to be a monomorphism.

Theorem 1.8. Let U and V have the same finite dimension n. A linear transformation of(J into V is an isomorphism if and only if it is an epimorphism. c is an isomorphism if and only if it is a monomorphism.

PROOF. It is part of the definition of an isomorphism that it is both an epimorphism and a monomorphism. Suppose 6 is an epimorphism. p(c) — n and v(c) = O by Theorem 1.6. Hence, is a monomorphism. Conversely if 6 is a monomorphism, then p(c) = O and, by Theorem 1.6, p(c) = n. Hence, c is an epimorphism. 

Thus a linear transformation c of U into V is an isomorphism if two of the following three conditions are satisfied: (l) dim U = dim V, (2) c is an epimorphism, (3) c is a monomorphism.

Theorem 1.9. p(T) = p(T0) + dim {Im(o) KG)}.

PROOF. Let T' be a new linear transformation defined on Im(c) mapping Im(c) into W so that for all e Im(c), T'(u) = T(u). Then K(T') = 1m(Ü) n K(T) and p(T') = dim = dim TÜ(U) = p(TC). Then Theorem 1.6 takes the form p(T') + p(T') dim Im(c),

or p(TÜ) + dim {Im(o) n K(T)} = p(c). 

Corollary 1.10. p(TÜ) = dim {Im(c) +

PROOF. This follows from Theorem 1.9 by application of Theorem 4.8 of Chapter I. 

Corollary 1.11. IfK(T) c Im@), then p(c)

Theorem 1.12. The rank of a product of linear transformations is less than or equal to the rank of either factor: p(TÜ) min {p(T), p(c)}.

PROOF. The rank of TO is the dimension c T(V). Thus considering dim c(U) as the "n" and dim T(V) as the "m" of Theorem 1.3 we see that dim TC(U) = p(TÜ) min {dim c(V), dim T (V)} = min {p(6), p(T)}. [3

Theorem 1.13. If is an epimorphism, then p(TÜ) = p(T). If T is a monomorphism, then p(TÜ) = p(o).

PROOF. If c is an epimorphism, then K(T) c Im(c) = V and Corollary 1.11 applies. Thus p(TÜ) = p(c) — 'V(T) = m — p(T) = p(T). If T is a monomorphism, then K(T) = {0} c Im(6) and Corollary 1.11 applies. Thus p(TÜ) = p(Ü) —

Corollary 1.14. The rank of a linear transformation is not changed by multiplication by an isomorphism (on either side). 

Theorem 1.15. is an epimorphism if and only if = 0 implies T = O. T is a monomorphism if and only if TC = O implies = 0.

PROOF. Suppose c is an epimorphism. Assume is defined and = O. If T # 0, there is a e V such that T(ß) O. Since is an epimorphism, there is an e U such that c(æ) = p. Then TO(æ) = T(ß) O. This is a contradiction and hence T = O. Now, suppose = 0 implies T = 0. If c is not an epimorphism then Im(c) is a subspace of V but Im(c) V. Let } be a basis of Im(ü), and extend this independent set to a basis

} of V. Define T(ßi) = for i > r and T(ßi) = O for

i r. Then = O and T O. This is a contradiction and, hence, c is an epimorphism.

Now, assume is defined and = O. Suppose T is a monomorphism. If c 0, there is an e U such that c(oc) O. Since T is a monomorphism, TO(oc) O. This is a contradiction and, hence, c = 0. Now assume = O implies c = O. If is not a monomorphism there is an u e U such that oc O and T(oc) = O. Let {0%, , . . , tin} be any basis of U. Define c(æ,.) = oc for each

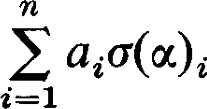
i. Then i ) = T(æ) = O for all i and = O. This is a contradiction and, hence, T is a monomorphism. C]

Corollary 1.16. c is an epimorphism if and only if TIC = implies T2. T is a monomorphism if and only if = implies — 

The statement that TIC = implies Tl = T2 is called a right-cancellation, and the statement that TCI = implies = is called a left-cancellation. Thus, an epimorphism is a linear transformation that can be cancelled on the right, and a monomorphism is a linear transformation that can be cancelled on the left.

Theorem 1.17. Let A = {011, . . . , n} be any basis of U. Let B = {61, ßn} be any n vectors in V (not necessarily linearly independent). There exists a uniquely determined linear transformation c of U into V such that c(oq.) = for i  n.

PROOF. Since A is a basis of U, any vector e U can be expressed uniquely in the form am . If is to be linear we must have

 c(æ) = = aißi G U.

It is a simple matter to verify that the mapping so defined is linear. 

Corollary 1.18. Let C = {71' . . . , yr} be any linearly independent set in U, where U is finite dimensional. Let D = {DI, Dr} be any r vectors in V. There exists a linear transformation of U into V such that c(Yi) = öi for



PROOF. Extend C to a basis of U. Define c(yt.) = for and define the values of on the other elements of the basis arbitrarily. This will yield a linear transformation with the desired properties. 

It should be clear that, if C is not already a basis, there are many ways to define c. It is worth pointing out that the independence of the set C is crucial to proving the existence of the linear transformation with the desired properties. Otherwise, a linear relation among the elements of C would impose a corresponding linear relation among the elements of D, which would mean that D could not be arbitrary.

Theorem 1.17 establishes, for one thing, that linear transformations really do exist. Moreover, they exist in abundance. The real utility of this theorem and its corollary is that it enables us to establish the existence of a linear transformation with some desirable property with great convenience. All we have to do is to define this function on an independent set.

Definition. A linear transformation of V into itself with the property that  — is called a projection.

Theorem 1.19. If7T is a projection of V into itself, then V = Im(T) 9 K(T) and acts like the identity on Im(T).

PROOF. For e V, let = T(æ). Then — 7T2(æ) = ff(æ) = ml. This shows that acts like the identity on Im(T). Let — — 0'1. Then 7T(u2) — T(oc) = o. Thus = + where e Im(ff) and e Clearly, 1m(T) n K(T) =

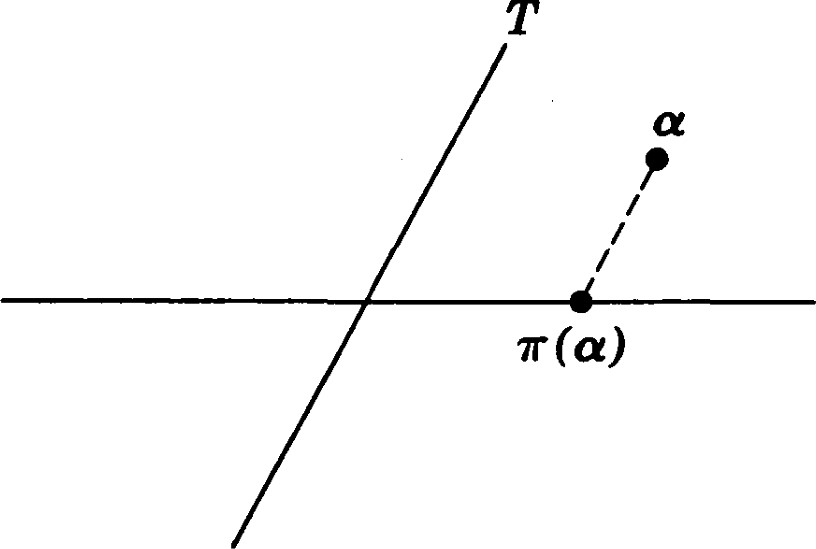
s

Fig. 1

If S = Im(T) and T = KG), we say that T is a projection of V onto S along T. In the case where V is the real plane, Fig. I indicates the interpretation of these words. a. is projected onto a point of S in a direction parallel to T.

EXERCISES

1. Show that x2)) (x2, $1) defines a linear transformation of R2 into itself.
2. Let $2)) —Xl) and $2)) (Xl, —CD. Determine + 02, and 0201.
3. Let U  and let x  0) where k < n. Describe Im(c) and K(o).
4. Let 4)) = (3X1 — 2X2 — — 4X4, + — 34). Show that c is a linear transformation. Determine the kernel of c.



—

1. Let 4, 4)) = (2$1 + + — Find

a basis of c(U). (Hint: Take particular values of the to find a spanning set for Ü(u).) Find a basis of K(c).

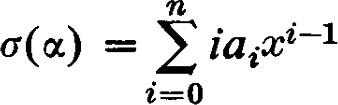
* 1. Let D denote the operator of differentiation,

dy d2Y

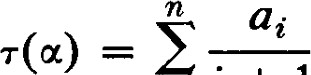
, D2(y) = — etc.

Show that Dn is a linear transformation, and also that p(D) is a linear transformation if p(D) is a polynomial in D with constant coeffcients. (Here we must assume that the space of functions on which D is defined contains only functions differentiable at least as often as the degree of p(D).)

* 1. Let U = V and let c and T be linear transformations of U into itself. In this case and are both defined. Construct an example to show that it is not always true that
  2. Let U = V = P, the space of polynomials in with coemcients in R. For O aiXi let



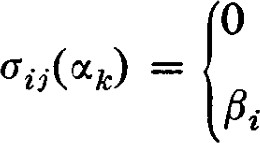
 and

xt+l

i=0i 1

 Show that = 1, but that 1.

* 1. Show that if two scalar transformations coincide on U then the defining scalars are equal.
  2. Let be a linear transformation of U into V and let A = {a} be a basis of U. Show that if the values {a(æ ) c(ocn)} are known, then the value of ff(u) can be computed for each e U.
  3. Let U and V be vector spaces of dimensions n and m, respectively, over the same field F. We have already commented that the set of all linear transformations of U into V forms a vector space. Give the details of the proof of this assertion. Let A = {al, , an} be a basis of U and B = {#1, , pm} be a basis of V. Let q, be the linear transformation of U into V such that

O if k if k = j.

Show that {oz.,. 1 n} is a basis of this vector space.

For the following sequence of problems let dim U = n and dim V = m. Let be a linear transformation of U into V and T a linear transformation of V into W.

12. Show that p(o) p(TÜ) + v(T). (Hint: Let V' = o(U) and apply Theorem

1.6 to T defined on V'.)

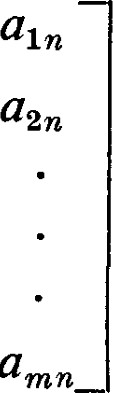
* 1. Show that max {O, PG) + p(T) — m} p(Tff) min {p(T), p(o)}.



* 1. Show that max {n — m + PG), v(o)} v(TÜ) min {n, + PG)}. (For m = n this inequality is known as Sylvester's law of nullity.)
  2. Show that if PG) = O, then p(TÜ) = p(c).
  3. It is not generally true that v(c) = O implies P(TC) = p(T). Construct an example to illustrate this fact. (Hint: Let m be very large.)
  4. Show that if m = n and v(a) = O, then p(TC) = PG).
  5. Show that if and are linear transformations of U into V, then p(C1 + 62) min {m, n, p&l) + p(C2)}•
  6. Show that I p&l) —  + 02).
  7. If S is any subspace of V there is a subspace T such that V = S T. Then every e V can be represented uniquely in the form = + where e S and e T. Show that the mapping which maps onto is a linear transformation. Show that T is the kernel of T. Show that 72 = T. The mapping is called a projection of V onto S along T.
  8. (Continuation) Let be a projection. Show that 1 — T is also a projection. What is the kernel of 1 — r? Onto what subspace is 1 — a projection? Show that 7T(1 — T) = O.

# 2 1 Matrices

Definition. A matrix over a field F is a rectangular array of scalars. The array will be written in the form

all a12 a21 a22

(2.1)

aml am2

whenever we wish to display all the elements in the array or show the form of the array. A matrix with m rows and n columns is called an m x n matrix. An n x n matrix is said to be of order n.

We often abbreviate a matrix written in the form above to [aijJ where the first index denotes the number of the row and the second index denotes the number of the column. The particular letter appearing in each index position is immaterial; it is the position that is important. With this convention atj is a scalar and [aij] is a matrix. Whereas the elements aii and an need not be equal, we consider the matrices and [an] to be identical since both and [an] stand for the entire matrix. As a further convenience we often use upper case Latin italic letters to denote matrices; A = [au]. Whenever we use lower case Latin italic letters to denote the scalars appearing

 Il.

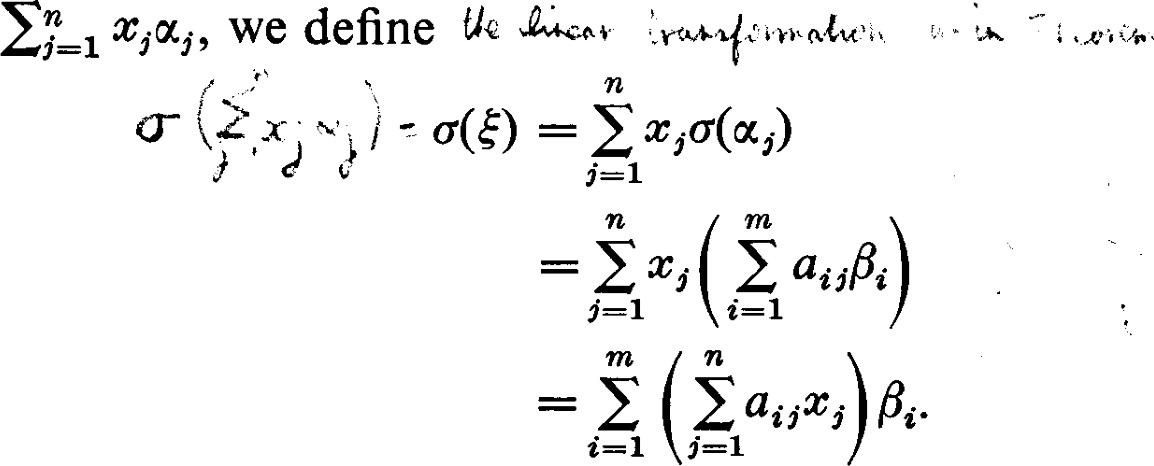
in the matrix, we use the corresponding upper case Latin italic letter to denote the matrix. The matrix in which all scalars are zero is denoted by 0 (the third use of this symbol!). The at, appearing in the array [at,] are called the elements of Two matrices are equal if and only if they have exactly the same elements. The main diagonal of the matrix is the set of elements {all, . , att} where t = min {m, n}. A diagonal matrix is a square matrix in which the elements not in the main diagonal are zero.

Matrices can be used to represent a variety of different mathematical concepts. The way matrices are manipulated depends on the objects which they represent. Considering the wide variety of situations in which matrices have found application, there is a remarkable similarity in the operations performed on matrices in these situations. There are differences too, however, and to understand these differences we must understand the object represented and what information can be expected by manipulating with the matrices. We first investigate the properties of matrices as representations of linear transformations. Not only do the matrices provide us with a convenient means of doing whatever computation is necessary with linear transformations, but the theory of vector spaces and linear transformations also proves to be a powerful tool in developing the properties of matrices.

Let U be a vector space of dimension n and V a vector space of dimension m, both over the same field F. Let A = {oc , un} be an arbitrary but fixed basis of U, and let B = {61, 18 } be an arbitrary but fixed basis of V. Let c be a linear transformation of U into V. Since c(ocj) e V, can be expressed uniquely as a linear combination of the elements of B •

= aißt• (2.2)

We define the matrix representing with respect to the bases A and B to be the matrix A = [at.j].

The correspondence between linear transformations and matrices is actually one-to-one and onto. Given the linear transformation c, the at, exist because B spans V, and they are unique because B is linearly independent. On the other hand, let A = [at,] be any m X n matrix. We can define = ajjßi for each e A, and then we can extend the proposed linear transformation to all of (J by the condition that it be linear. Thus, if =

(2.3)



can be extended to all of U because A spans U, and the result is well defined (unique) because A is linearly independent.

Here are some examples of linear transformations and the matrices which represent them. Consider the real plane R2 = U = V. Let A = B 

(0, l)}. A 900 rotation counterclockwise would send (l, O) onto (0, l) and it would send (0, l) onto (—1, 0). Since 0)) = O . (1, 0) + 1 (0, 1) and

l)) 1) • (l, 0) + 0 • (0, l), c is represented by the matrix

I 0 

The elements appearing in a column are the coordinates of each image of a basis vector under a transformation. 

In general, a rotation counterclockwise through an angle of 0 will send (1, O) onto (cos 0, sin 0) and (0, l) onto ( —sin 0, cos 0). Thus this rotation is represented by

cos 0 —sin 0

(2.4) sin 0 cos 0 

Suppose now that T is another linear transformation of U into V represented by the matrix B =  Then for the sum + T we have

(O +  = + T(æj) = aißi + bißi

= + bli)ßt. (2.5)

Thus + T is represented by the matrix [ai,. + bij]. Accordingly, we define the sum of two matrices to be that matrix obtained by the addition of the corresponding elements in the two arrays; A + B = [au + bi,.l is the matrix corresponding to c + T. The sum of two matrices is defined if and only if the two matrices have the same number of rows and the same number of columns.

If a is any scalar, for the linear transformation ac we have

= a aiißi = (aati)ßi. (2.6)

Thus ac is represented by the matrix [aac.,.]. We therefore define scalar multiplication by the rule aA = [aat.,.].

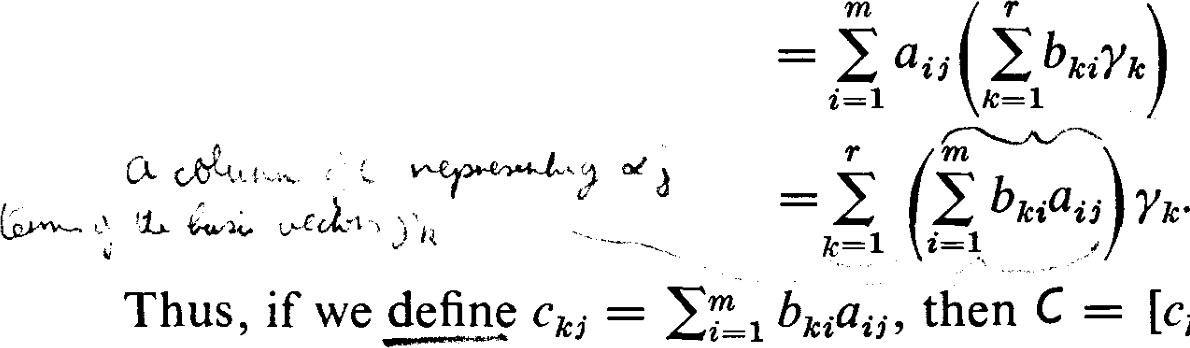
Let W be a third vector space of dimension r over the field F, and let , Yr} be an arbitrary but fixed basis of W. If the linear transformation c of U into Vis represented by the m x n matrix A = [at,] and the

linear transformation T of V into W is represented by the r x m matrix B = [hi], what matrix represents the linear transformation TO of U into W?

= T ( außi

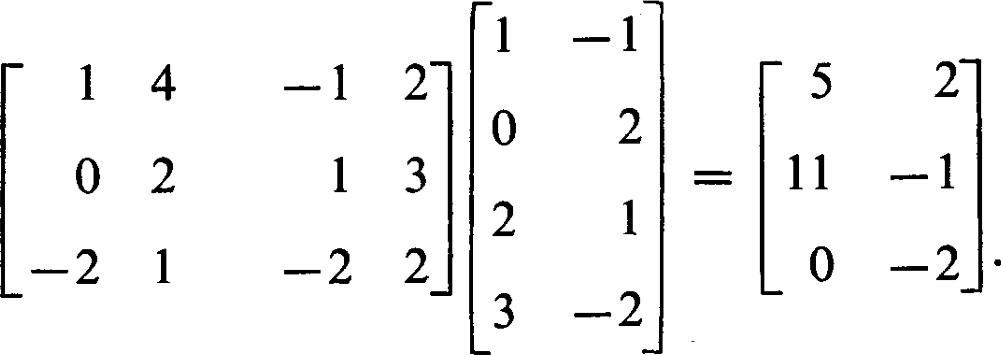
= ai5T(ßi)

 (2.7)

1 bicßii, then C = [Ck,.] is the matrix representing the product transformation 76. Accordingly, we call C the matrix product of B and A, in that order: C = BA.

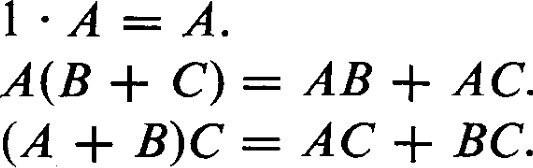
For computational purposes it is customary to write the arrays of B and A side by side. The element of the product is then obtained by multiplying the corresponding elements of row k of B and column j of A and adding. We can trace the elements of row k of B with a finger of the left hand while at the same time tracing the elements of column j of A with a finger of the right hand. At each step we compute the product of the corresponding elements and accumulate the sum as we go along. Using this simple rule we can, with practice, become quite proficient, even to the point of doing "without hands."

Check the process in the following examples:



All definitions and properties we have established for linear transformations can be carried over immediately for matrices. For example, we have:

1. O • A = 0. (The "0" on the left is a scalar, the "0" on the right is a matrix with the same number of rows and columns as A.)

3. 

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

Of course, in each of the above statements we must assume the operations proposed are well defined. For example, in 3, B and C must be the same size and A must have the same number of columns as B and C have rows.

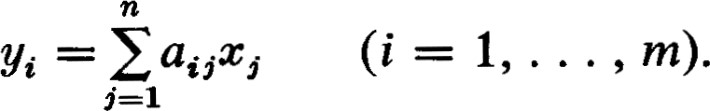
The rank and nullity of a matrix A are the rank and nullity of the associated linear transformation, respectively.

Theorem 2.1. For an m X n matrix A, the rank of A plus the nullity of A is equal to n. The rank of a product BA is less than or equal to the rank of either factor.

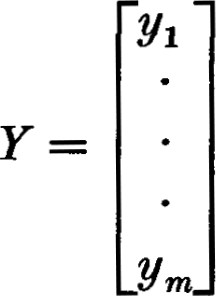
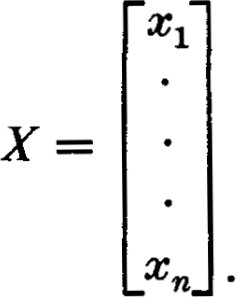
These statements have been established for linear transformations and therefore hold for their corresponding matrices. C]

The rank of c is the dimension of the subspace Im(c) of V. Since Im(c) is spanned by {c(æl), c(æn)}, p(c) is the number of elements in a maximal linearly independent subset of {CGI), c(æn)}. Expressed in terms of coordinates, c(oc,.) = atißi is represented by the m-tuple a ) which is the m-tuple in column j of the matrix Thus p(c) = p(A) is also equal to the maximum number of linearly independent columns of A. This is usually called the column rank of a matrix A, and the maximum number of linearly independent rows of A is called the row rank of A. We, however, show before long that the number of linearly independent rows in a matrix is equal to the number of linearly independent columns. Until that time we consider "rank" and "column rank" as synonymous.

Returning to Equation (2.3), we see that, if e U is represented by ($1, x ) and the linear transformation c of U into V is represented by the matrix A = then e V is represented by (VI, . . . , ym) where

(2.8)

In view of the definition of matrix multiplication given by Equation (2.7) we can interpret Equations (2.8) as a matrix product of the form

Y = AX (2.9) where  and

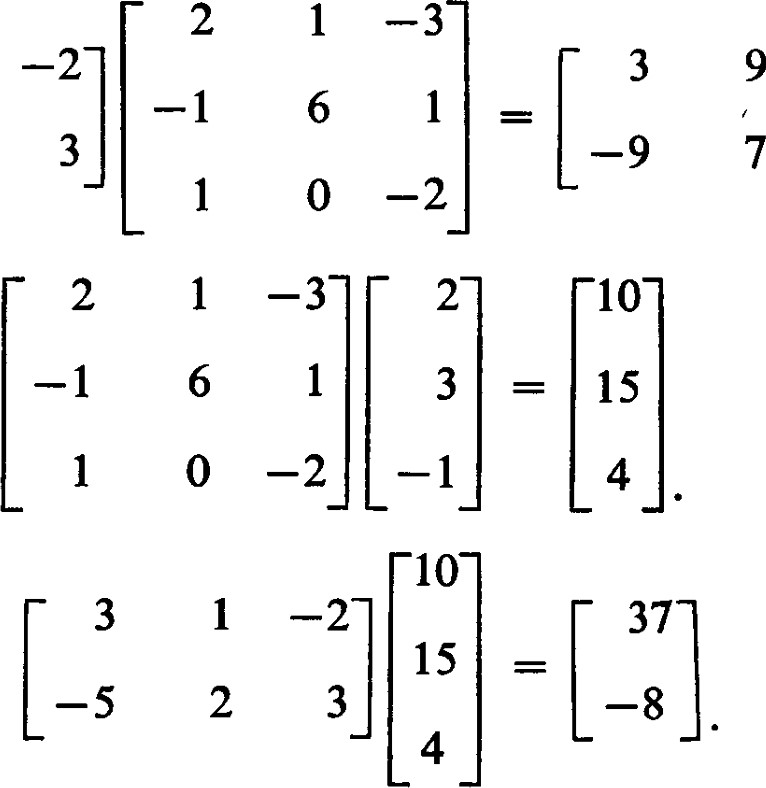
This single matric equation contains the m equations in (2.8).

We have already used the n-tuple (c , cn) to represent the vector c . Because of the usefulness of equation (2.9) we also find it convenient to represent by the one-column matrix X. In fact, since it is somewhat wasteful of space and otherwise awkward to display one-column matrices we use the n-tuple ($1, , cn) to represent not only the vector but also the column matrix X. With this convention [Xl • • • cn] is a one-row matrix and @ , cn) is a one-column matrix.

Notice that we have now used matrices for two different purposes, (1) to represent linear transformations, and (2) to represent vectors. The single matric equation Y = AX contains some matrices used in each way.

EXERCISES

1. Verify the matrix multiplication in the following examples :

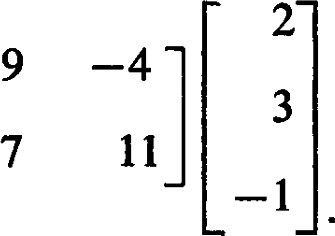
(a)

11

(b)

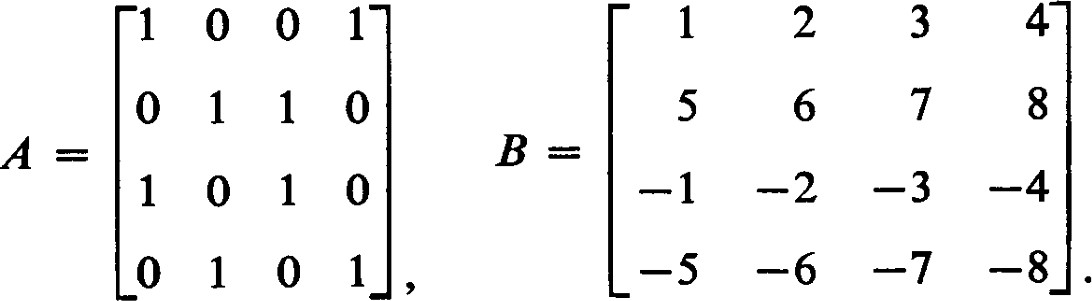
# (c)

1. Compute



Interpret the answer to this problem in terms of the computations in Exercise 1.

1. Find AB and BA if



1. Let o be a linear transformation of R2 into itself that maps (1 , O) onto (3, — and (O, 1) onto (—1, 2). Determine the matrix representing with respect to the bases A — B = 0), (0, 1)}.
2. Let be a linear transformation of R2 into itself that maps (1, 1) onto (2, —3) and (1, —1) onto (4, —7). Determine the matrix representing with respect to the bases A = B = {(1, 0), (O, 1)}. (Hint: We must determine the effect of when it is applied to (1, O) and (0, 1). Use the fact that (1, O) = 1) + —1) and the linearity of o.)
3. It happens that the linear transformation defined in Exercise 4 is one-to-one, that is, does not map two different vectors onto the same vector. Thus, there is a linear transformation that maps (3, —1) onto (1, O) and (—1, 2) onto (O, 1). This linear transformation reverses the mapping given by c. Determine the matrix representing it with respect to the same bases.
4. Let us consider the geometric meaning of linear transformations. A linear transformation of R2 into itself leaves the origin fixed (why?) and maps straight lines into straight lines. (The word "into" is required here because the image of a straight line may be another straight line or it may be a single point.) Prove that the image of a straight line is a subset of a straight line. (Hint: Let be represented by the matrix

all a12 a21 acn 

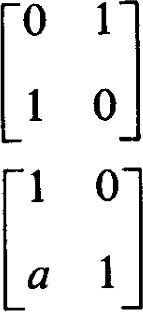
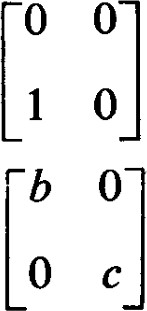
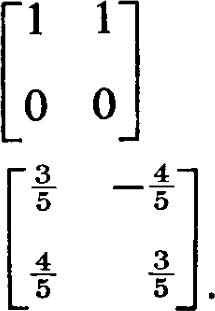
Then maps (c, y) onto (anx + auy, anc + a,uy). Now show that if @, y) satisfies the equation ax + by = c its image satisfies the equation

(aa22 — ba21)x + (allb — al'2a)Y = (alla22 — a12a21)c•)

1. (Continuation) We say that a straight line is mapped onto itself if every point on the line is mapped onto a point on the line (but not all onto the same point) even though the points on the line may be moved around.
2. A linear transformation maps (1, 0) onto (—1, O) and (O, 1) onto (O, —1). Show that every line through the origin is mapped onto itself. Show that each such line is mapped onto itself with the sense of direction inverted. This linear transformation is called an inversion with respect to the origin. Find the matrix representing this linear transformation with respect to the basis {(1, O), (O, 1)}.
3. A linear transformation maps (1, 1) onto (—1, —1) and leaves (1 fixed. Show that every line perpendicular to the line + = O is mapped onto itself with the sense of direction inverted. Show that every point on the line = O is left fixed. Which lines through the origin are mapped onto themselves? This linear transformation is called a reflection about the line + — O. Find the matrix representing this linear transformation with respect to the basis {(1, O), (O, 1)}. Find the matrix representing this linear transformation with respect to the basis {(1, 1) (1
4. A liner transformation maps (1, 1) onto (2, 2) and (1, —1) onto (3

Show that the lines through the origin and passing through the points (1, 1) and (1, —1) are mapped onto themselves and that no other lines are mapped onto themselves. Find the matrices representing this linear transformation with respect to the bases {(1, O), (O, 1)} and {(1 1) (1

1. A linear transformation leaves (1, 0) fixed and maps (O, 1) onto (1, 1). Show that each line = c is mapped onto itself and translated within itself a distance equal to c. This linear transformation is called a shear. Which lines through the origin are mapped onto themselves? Find the matrix representing this linear transformation with respect to the basis {(1 , O), (O, 1)}.
2. A linear transformation maps (1, O) onto (IS, 11 3) and (O, 1) onto (—1 2 Show that every line through the origin is rotated counterclockwise through the angle O = arc cos 13. This linear transformation is called a rotation. Find the matrix representing this linear transformation with respect to the basis {(1, O),
3. A linear transformation maps (1 , O) onto (å, { ) and (0, 1) onto å, å). Show that each point on the line 2x1 + = 3c is mapped onto the single point (c, c). The line = 0 is left fixed. The only other line through the origin which is mapped into itself is the line 2x1 + = O. This linear transformation is called a projection onto the line 2 = 0 parallel to the line 2x1 + - — O. Find the matrices representing this linear transformation with respect to the bases {(1 , O), (0, 1)) and {(1 1) (1
4. (Continuation) Describe the geometric effect of each of the linear transformations of R2 into itself represented by the matrices

(a)(b)(c)

(d)(f)

(Hint: In Exercise 7 we have shown that straight lines are mapped into straight lines. We already know that linear transformations map the origin onto the origin. Thus it is relatively easy to determine what happens to straight lines passing through the origin. For example, to see what happens to the Cl-axis it is suffcient to see what happens to the point (1, 0). Among the transformations given appear a rotation, a reflection, two projections, and one shear.)

1. (Continuation) For the linear transformations given in Exercise 9 find all lines through the origin which are mapped onto or into themselves.
2. Let U = R2 and V = R3 and o be a linear transformation of U into V that maps (1, 1) onto (0, 1, 2) and (—1, 1) onto (2, 1, 0). Determine the matrix that represents with respect to the bases A = {(1, 0), (0, 1)} in B = {(1, O, O), (O, 1, O), (0, O, 1)} in R3 . (Hint: 1) 1 ( -1, 1) = (1,
3. What is the effect of multiplying an n x n matrix A by an n x n diagonal matrix D? What is the difference between AD and DA ?
4. Let a and b be two numbers such that a b. Find all 2 x 2 matrices A such that
5. Show that the matrix C = [ac•bj] has rank one if not all ai and not all bi are zero. (Hint: Use Theorem 1.12.)
6. Let a, b, c, and d be given numbers (real or complex) and consider the function ax + b cc + d

Let g be another function of the same form. Show that "where gf@) = is a function that can also be written in the same form. Show that each of these functions can be represented by a matrix in such a way that the matrix representing g/is the product of the matrices representing g andf. Show that the inverse function exists if and only if ad — bc O. To what does the function reduce if ad — bc = O?

1. Consider complex numbers of the form + yi (where c and y are real numbers and 1 • 2 = —1) and represent such a complex number by the duple @, y) in R2 . Let a + bi be a fixed complex number. Consider the function f defined by the rule

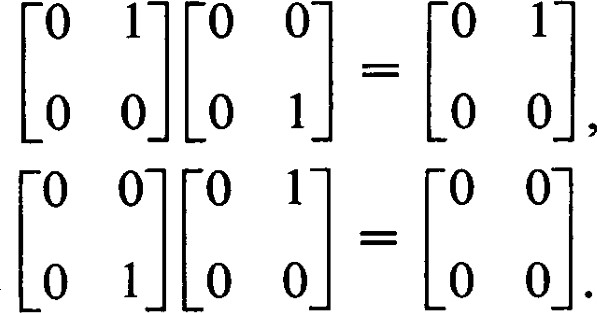
= u + vi.

1. Show that this function is a linear transformation of R2 into itself mapping @, y) onto (u, v).
2. Find the matrix representing this linear transformation with respect to the basis {(1, O), (O, 1)}.
3. Find the matrix which represents the linear transformation obtained by using c + di in place of a + bi. Compute the product of these two matrices. Do they commute ?
4. Determine the complex number which can be used in place of a + bi to obtain a transformation represented by this matrix product. How is this complex number related to a + bi and c + di?

17. Show by example that it is possible for two matrices A and B to have the same rank while A2 and B2 have different ranks.

3 1 Non-singular Matrices

Let us consider the case where U = V, that is, we are considering transformations of V into itself. Generally, a homomorphism of a set into itself is called an endomorphism. We consider a fixed basis in V and represent the linear transformation of V into itself with respect to that basis. In this case the matrices are square or n X n matrices. Since the transformations we are considering map V into itself any finite number of them can be iterated in any order. The commutative law does not hold, however. The same remarks hold for square matrices. They can be multiplied in any order but the commutative law does not hold. For example



The linear transformation that leaves every element of V fixed is the identity transformation. We denote the identity transformation by 1 , the scalar identity. Clearly, the identity transformation is represented by the matrix I = [öi,.] for any choice of the basis. Notice that IA = Al = A for any n X n matrix A. I is called the identity matrix, or unit matrix, of order n. If we wish to point out the dimension of the space we write In for the identity matrix of order n. The scalar transformation a is represented by the matrix al. Matrices of the form al are called scalar matrices.

Definition. A one-to-one linear transformation o of a vector space onto itself is called an automorphism. An automorphism is only a special kind of isomorphism for which the domain and codomain are the same space. If c(oc) = a, the mapping = oc is called the inverse transformation of c. The rotations represented in Section 2 are examples of automorphisms.

  Theorem 3.1. The inverse of an automorphism c is an automorphism.

Theorem 3.2 A linear transformation T of an n-dimensional vector space into itself is an automorphism if and only if it is of rank n; that is, if and only if it is an epimorphism.

•Theorem 3.3. A linear transformation c of an n-dimensional vector space into itself is an automorphism if and only if its nullity is 0, that is, if and only if it is a monomorphism.

PROOF (of Theorems 3.1, 3.2, and 3.3). These properties have already been established for isomorphisms. 

Since it is clear that transformations of rank less than n do not have •inverses because they are not onto, we see that automorphisms are the ' only linear transformations which have inverses. A linear transformation that has an inverse is said to be non-singular or invertible; otherwise it is said to be singular. Let A be the matrix representing the automorphism c, and let be the matrix representing the inverse transformation a—I. The matrix A—IA represents the transformation TIC. Since c—lc is the identity transformation, we must have A—IA = I. But c is also the inverse transformation of so that cc—I = 1 and AA —I = I. We shall refer to as the inverse of A. A matrix that has an inverse is said to be nonsingular or invertible. Only a square matrix can have an inverse.

On the other hand suppose that for the matrix A there exists a matrix B such that BA = I. Since I is of rank n, A must also be of rank n and, therefore, A represents an automorphism c. Furthermore, the linear transformation which B represents is necessarily the inverse transformation  since the product with must yield the identity transformation. Thus B = A—I . The same kind of argument shows that if C is a matrix such that AC = I, then C = A -I . Thus we have shown:

Theorem 3.4. If A and B are square matrices such that BA = I, then AB = I. If A and B are square matrices such that AB = I, then BA = I. In either case B is the unique inverse of A. 

Theorem 3.5. If A and B are non-singular, then (l) AB is non-singular and

 = (2) is non-singular and = A, (3) for a 0, aA is non-singular and

PROOF. In view of the remarks preceding Theorem 3.4 it is sumcient in each case to produce a matrix which will act as a left inverse.

1. = B-11B = B-lB = 1.
2. AA-I = 1.

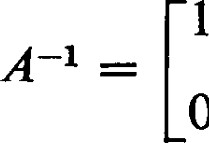


Theorem 3.6. If A is non-singular, we can solve uniquely the equations XA = B and A Y = Bfor any matrix B of the proper size, but the two solutions need not be equal.

PROOF. Solutions exist since =  = B and  = = B. The solutions aré unique since for any C having the property that CA = B we have C = CAA -I = BA—I , and similarly with any solution



As an example illustrating the last statement of the theorem, let

 1 2 1

|  |  |  |
| --- | --- | --- |
| Then |  |  |
| X = BA -I | and | Y = A -IB |

2 1

We add the remark that for non-singular A, the solution of XA = B exists and is unique if B has n columns, and the solution of AY = B exists and is unique if B has n rows. The proof given for Theorem 3.6 applies without change.

Theorem 3.7. The rank of a (not necessarily square) matrix is not changed by multiplication by a non-singular matrix.

PROOF. Let A be non-singular and let B be of rank p. Then by Theorem

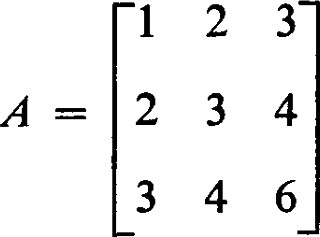
2.1 AB is of rank r p, and A -I (AB) = B is of rank p r. Thus r The proof that BA is of rank p is similar. 

Theorem l . 14 states the corresponding property for linear transformations. The existence or non-existence of the inverse of a square matrix depends on the matrix itself and not on whether it represents a linear transformation of a vector space into itself or a linear transformation of one vector space into another. Thus it is convenient and consistent to extend our usage of the term "non-singular" to include isomorphisms. Accordingly any square matrix with an inverse is non-singular.

Let U and V be vector spaces of dimension n over the field F. Let A = {11, , be a basis of U and B ßn} be a basis of V. If  is any vector in U we can define to be 1 Xißi. It is easily seen that is an isomorphism and that st and c(é) are both repre sented by ($1, xn) e Fn . Thus any two vector spaces of the same dimension over F are isomorphic. As far as their internal structure is concerned they are indistinguishable. Whatever properties may serve to distinguish them are, by definition, not vector space properties.

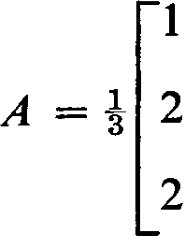
EXERCISES

1. Show that the inverse of

—2 o 1 o 3 -2

1 —2 1

1. Find the square of the matrix

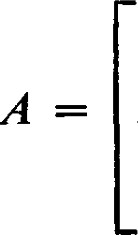
 2 2

-2 1

1 --2

What is the inverse of A? (Geometrically, this matrix represents a 1800 rotation about the line containing the vector (2, 1, 1). The inverse obtained is therefore not surprising.)

1. Compute the image of the vector (1, —2, 1) under the linear transformation represented by the matrix

1 2 3 2 3 4 o 1 2

Show that A cannot have an inverse.

1. Since

11 12 3 -1

21 $22 —5 2

3 -1

3C1111 

3X21 — 5X2221 + 242

we can find the inverse of by solving the equations

—5 2

3C11 — 5C12

11 + 2$12

## 3X21

—X21 2X22 = 1.

Solve these equations and check your answer by showing that this gives the inverse matrix.

We have not as yet developed convenient and effective methods for obtaining the inverse of a given matrix. Such methods are developed later in this chapter and in the following chapter. If we know the geometric meaning of the matrix, however, it is often possible to obtain the inverse with very little work.

3 4

 5. The matrix 4 3 represents a rotation about the origin through the angle

1. = arc cos s. What rotation would be the inverse of this rotation? What matrix would represent this inverse rotation? Show that this matrix is the inverse of the given matrix.

6. The matrix represents a reflection about the line + — —1 

What operation is the inverse of this reflection? What matrix represents the inverse operation? Show that this matrix is the inverse of the given matrix.

1. 1
2. The matrix represents a shear. The inverse transformation is also a

shear. Which one? What matrix represents the inverse shear? Show that this matrix is the inverse of the given matrix.

1. Show that the transformation that maps ($1, $2, c3) onto ($3, —$1, $2) is an automorphism of F3. Find the matrix representing this automorphism and its inverse with respect to the basis {(1, O, O), (O, 1, O), (O, O, 1)}.
2. Show that an automorphism of a vector space maps every subspace onto a subspace of the same dimension.
3. Find an example to show that there exist non-square matrices A and B such that AB = I. Specifically, show that there is an m x n matrix A and an n x m matrix B such that AB is the m x m identity. Show that BA is not the n x n identity. Prove in general that if m n, then AB and BA cannot both be identity matrices.

4 1 Change of Basis

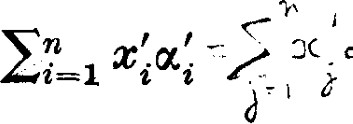
We have represented vectors and linear transformations as n-tuples and matrices with respect to arbitrary but fixed bases. A very natural question arises: What changes occur in these representations if other choices for bases are made? The vectors and linear transformations have meaning independent of any particular choice of bases, independent of any coordinate systems, but their representations are entirely dependent on the bases chosen.

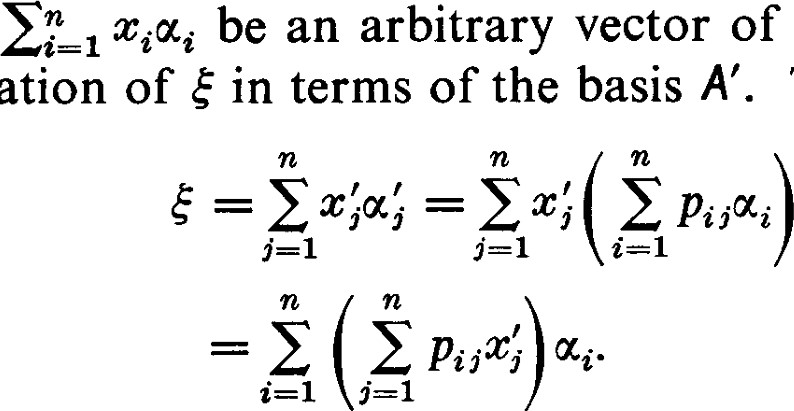
Definition. Let A = {al, . . . , G} and A u'} be bases of the vector space U. In a typical "change of basis" situation the representations of various vectors and linear transformations are known in terms of the basis A, and we wish to determine their representations in terms of the basis A'. In this connection, we refer to A as the "old" basis and to A' as the "new" basis. Each is expressible as a linear combination of the elements of A; that is,

= Pii%. (4.1)

The associated matrix P = is called the matrix of transition from the basis A to the basis AV .

The columns of P are the n-tuples representing the new basis vectors in terms of the old basis. This simple observation is worth remembering as it is usually the key to determining P when a change of basis is made. Since the columns of P are the representations of the basis A' they are linearly independent and P has rank n. Thus P is non-singular.

Now let =U and let = be the representation of in terms of the basis A'. Then 



of

(4.2)

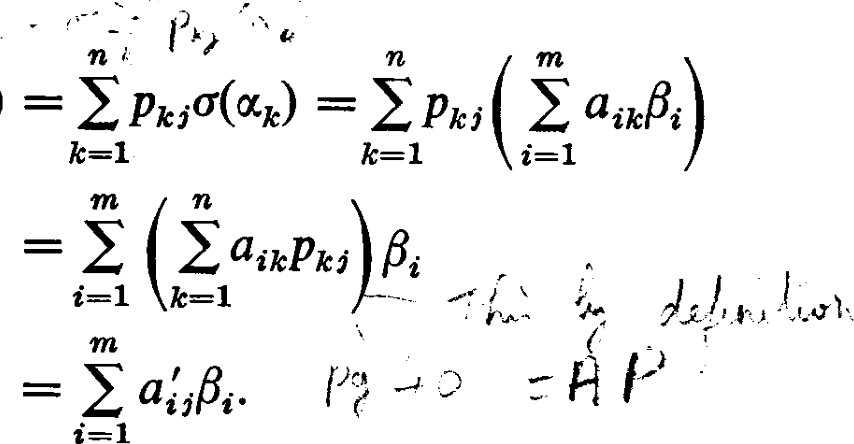
Since the representation of s: with respect to the basis A is unique we see that = 21=1 p„xj. Notice that the rows of P are used to express the old coordinates of in terms of the new coordinates. For emphasis and contradistinction, we repeat that the columns of P are used to express the new basis vectors in terms of the old basis vectors.

Let X = ($1, c ) and X' = @'1, c') ben x I matrices representing the vector with respect to the bases A and A'. Then the set of relations {Ci = 21=1 pijxj} can be written as the single matric equation

X = PX'. (4.3)

4 Change of Basis

Now suppose that we have a linear transformation c of U into V and that A = [at,] is the matrix representing c with respect to the bases A in U and B = {P} in V. We shall now determine the representation of c with respect to the bases A' and B.

c(æ'j)

(4.4)

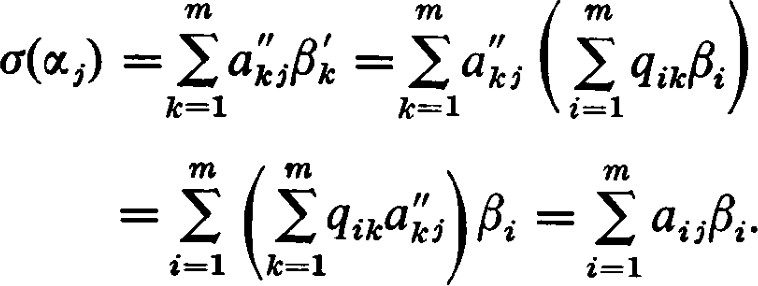
Since B is a basis, a;i — E?\_ 1 ai3cp,cj and the matrix A' = representing c with respect to the bases A' and B is related to A by the matric equation

A' = AP. (4.5)

This relation can also be demonstrated in a slightly different way. For an arbitrary = c e U let = YiÅi. Then we have AX = A(PX') = (AP)X'. (4.6)

Thus AP is a matrix representing c with respect to the bases A' and B. Since the matrix representing c is uniquely determined by the choice of bases we have A' = AP.

Now consider the effect of a change of basis in the image space V. Thus let B be replaced by the basis B , ß'm}. Let Q = [qij] be the matrix of transition from B to B', that is, ß'j = Etm\_l qijßi. Then if A" [a'i',.] represents c with respect to the bases A and B' we have

 (4.7)

Since the representation of c(ocj) in terms of the basis B is unique we see that A = QA", or

A" = Q -IA. (4.8)

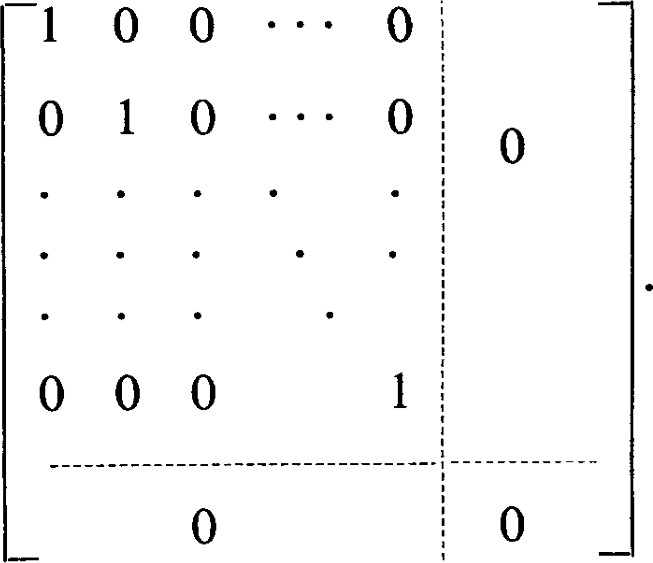
Combining these results, we see that, if both changes of bases are made at once, the new matrix representing c is Q-IAP.

As in the proof of Theorem 1.6 we can choose a new basis A' = {oc of U such that the last v = n — p basis elements form a basis of K(c). Since , is a basis of c(U) and is linearly independent in V, it can

be extended to a basis B' of V. With respect to the bases A' and B' we have

= ß'i for j p while = O for j > p. Thus the new matrix Q—IAP representing c is of the form

p columns v columns

p rows

m — p rows

Thus we have

Theorem 4.1. If A is any m x n matrix of rank p, there exist a nonsingular n X n matrix P and a non-singular m x m matrix Q such that A' = Q—IAP has the first p elements of the main diagonal equal to 1, and all other elements equal to zero. 

When A and B are unrestricted we can always obtain this relatively simple representation of a linear transformation by a proper choice of bases. More interesting situations occur when A and B are restricted. Suppose, for example, that we take U = V and A = B. In this case there is but one basis to change and but one matrix of transition, that is, P = Q. In this case it is not possible to obtain a form of the matrix representing c as simple as that obtained in Theorem 4.1. We say that any two matrices representing the same linear transformation of a vector space V into itself are similar. This is equivalent to saying that two matrices A and A' are similar if and only if there exists a non-singular matrix of transition P such that A' P-IAP. This case occupies much of our attention in Chapters Ill and V.

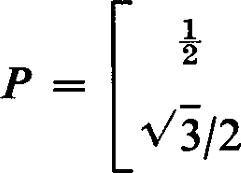
EXERCISES

1. In P3, the space of polynomials of degree 2 or smaller with coemcients in F, let A = {I x)

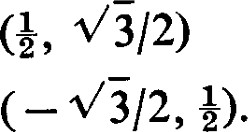
2 — — 2, P3(x) is also a basis. Find the matrix of transition from A to A'.

5 Hermite Normal Form

1. In many of the uses of the concepts of this section it is customary to take  Din)} as the old basis in Rn. Thus, in R2 let A — {(1, O), (O, 1)} and A' = Vi/2), (— s/j/2, l)}. Show that

 1

is the matrix of transition from A to A'.

1. (Continuation) With A' and A as in Exercise 2, find the matrix of transition R from A' to A. (Notice, in particular, that in Exercise 2 the columns of P are the components of the vectors in A' expressed in terms of basis A, whereas in this exercise the columns of R are the components of the vectors in A expressed in terms of the basis A'. Thus these two matrices of transition are determined relative to different bases.) Show that RP = I.
2. (Continuation) Consider the linear transformation of of R2 into itself which maps  onto  onto

Find the matrix A that represents with respect to the basis A.

You should obtain A = P. However, A and P do not represent the same thing. To see this, let = x2) be an arbitrary vector in R2 and compute by means of formula (2.9) and the new coordinates of by means of formula (4.3).

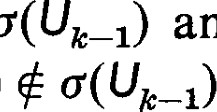
A little reflection will show that the results obtained are entirely reasonable. The matrix A represents a rotation of the real plane counterclockwise through an angle of 7/3. The matrix P represents a rotation of the coordinate axes counterclockwise through an angle of 7/3. In the latter case the motion of the plane relative to the coordinate axes is clockwise through an angle of 7/3.

1. In R3 let A = O, 0), (0, 1, 0), (0, O, 1)} and let A' = 1, 1), (1, O, 1),

(1, 1, O)}. Find the matrix of transition P from A to A' and the matrix of transition from A' to A.

1. Let A, B, and C be three bases of V. Let P be the matrix of transition from A to B and let Q be the matrix of transition from B to C. Is PQ or QP the matrix of transition from A to C? Compare the order of multiplication of matrices of transition and matrices representing linear transformation.
2. Use the results of Exercise 6 to resolve the question raised in the parenthetical remark of Exercise 3, and implicitly assumed in Exercise 5. If P is the matrix of transition from A to A' and Q is the matrix of transition from A' to A, show that

5 1 Hermite Normal Form

We may also ask how much simplification of the matrix representing a linear transformation c of U into V can be effected by a change of basis in V alone. Let A = VI, , an} be the given basis in U and let Uk = (0%, . . . , uk). The subspaces c(Uk) of V form a non-decreasing chain of subspaces with c c(Uk) and c(Un) = c(U). Since c(Uk) = o(lJlc 1) + we see from Theorem 4.8 of Chapter I that dim c(Uk) dim Ü(Ulc 1) + l ; that is, the dimensions of the c(Uk) do not increase by more than I at a time as k increases. Since dim c(Un) = p, the rank of c, an increase of exactly I must occur p times. For the other times, if any, we must have dim c(Uk) — dim and hence c(Uk) = c(lJk 1). We have an increase by I when c(otk) and no increase when 6(4) e 1).

Let kl, k2, . . . kp be those indices for which ka—l Let — Since ß'i ) , "'i\_1), the set , ß'p} is linearly independent (see Theorm 2.3, Chapter 1-2). Since {M, c(U) and c(U) is of dimension p, {M, . , F'} P is a basis of c(U). This set can be extended to a basis B' of V. Let us now determine the form of the matrix A' representing c with respect to the bases A and B'.

Since 0(4.) = ß't, column ki has a 1 in row i and all other elements of this column are O's. For ki < j < k c(æj) e c(Ukz) so that column j has o's below row i. In general, there is no restriction on the elements of column j in the first i rows. A' thus has the form

|  |  |
| --- | --- |
| column | column |

## 00 a

0 O O1 a '

2,k2+1

0 00 0

(5.1)

## 0 0 00 0

Once A and c are given, the ki and the set {F' , Å'p} are uniquely determined. There may be many ways to extend this set to the basis B', but the additional basis vectors do not affect the determination of A' since every element of c(U) can be expressed in terms of {F' F'} alone. Thus A' is uniquely determined by A and o.

Theorem 5.1. Given any m x n matrix A of rank p, there exists a nonsingular m X m matrix Q such that A' = Q—IA has the following form:

(1) There is at least one non-zero element in each of the first p rows of A' and the elements in all remaining rows are zero.

5 Hermite Normal Form

1. The first non-zero element appearing in row i (i p) is a I appearing in column ki, where kl < k2 < • • < k 
2. In column ki the only non-zero element is the I in row i.

The form A' is uniquely determined by A.

PROOF. In the applications of this theorem that we wish to make A is usually given alone without reference to any bases A and B, and often without reference to any linear transformation c. We can, however, introduce any two vector spaces U and V of dimensions n and m over F and let A be any basis of U and B be any basis of V. We can consider A as defining a linear transformation of U into V with respect to the bases A and B. The discussion preceding Theorem 5.1 shows that there is at least one non-singular matrix Q such that Q—IA satisfies conditions (l), (2), and (3).

Now suppose there are two non-singular matrices QI and Q2 such that Q—IA = A'l and Q—IA = A'2 both satisfy the conditions of the theorem. We wish to conclude that A'l = A'2. No matter how the vector spaces U and V are introduced and how the bases A and B are chosen we can regard QI and Q2 as matrices of transition in V. Thus A'l represents c with respect to bases A and B'l and A'2 represents c with respect to bases A and B;. But condition (3) says that for i p the ith basis element in both B'l and B'2 is c(oqc ). Thus the first p elements of B'l and B'2 are identical. Condition (1) says that the remaining basis elements have nothing to do with determining the coemcients in A'l and A'2. Thus A'l = A'2. 

We say that a matrix satisfying the conditions of Theorem 5.1 is in Hermite normal form. Often this form is called a row-echelon form. And sometimes the term, Hermite normal form, is reserved for a square matrix containing exactly the numbers that appear in the form we obtained in Theorem 5.1 with the change that row i beginning with a 1 in column ki is moved down to row kt. Thus each non-zero row begins on the main diagonal and each column with a I on the main diagonal is otherwise zero. In this text we have no particular need for this special form while the form described in Theorem 5.1 is one of the most useful tools at our disposal.

The usefulness of the Hermite normal form depends on its form, and the uniqueness of that form will enable us to develop effective and convenient short cuts for determining that form.

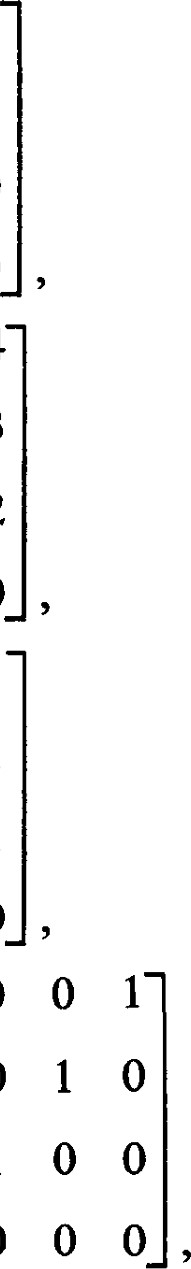
Definition. Given the matrix A, the matrix A T obtained from A by interchanging rows and columns in A is called the transpose of A. If A T = [a'i,.], the element a'ij appearing in row i column j of A T is the element aji appearing in row j column i of A. It is easy to show that (AB) T = BTA T. (See Exercise 4.)

Proposition 5.2. The number of linearly independent rows in a matrix is equal to the number of linearly independent columns.

PROOF. The number of linearly independent columns in a matrix A is its rank p. The Hermite normal form A' = Q—IA corresponding to A is also of rank p. For A' it is obvious that the number of linearly independent rows in A' is also equal to p, that is, the rank of (A') T is p. Since QT is nonsingular, the rank of A T = (QA') T = (A') TQT is also p. Thus the number of linearly independent rows in A is p. 

EXERCISES

1. Which of the following matrices are in Hermite normal form ?
   * + 1 o o 1



o

o

1

O

* + - o 1 o 1
  1. O o o 1 o o o O o o O o 2  4
     + 1 1 o 3
  2. o o o 1 2 o o o o o 1 O O o 1
     + 1 o o 1
  3. o o o 1 1 o O o O O o 1 o 1
     +  1 o
  4. o o  o

O o o o 1 o 1 o 1 o 1 1 o o o O o 1 o

O o O o 1

1. Determine the rank of each of the matrices given in Exercise 1.
2. Let and T be linear transformations mapping R3 into R2. Suppose that for a given pair of bases A for R3 and B for R2 , o and T are represented by

 1 1 0 1 0 1

and B —

# 0 0 1 0 1 0

respectively. Show that there is no basis B' of R2 such that B is the matrix representing c with respect to A and B'.

4. Show that

1. (A + = AT + BT,
2. (AB)T = BTA T

6 1 Elementary Operations and Elementary Matrices

Our purpose in this section is to develop convenient computational methods. We have been concerned with the representations of linear transformations by matrices and the changes these matrices undergo when a basis is changed. We now show that these changes can be effected by elementary operations on the rows and columns of the matrices.

We define three types of elementary operations on the rows of a matrix A.

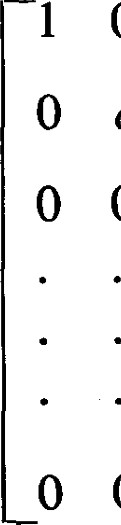
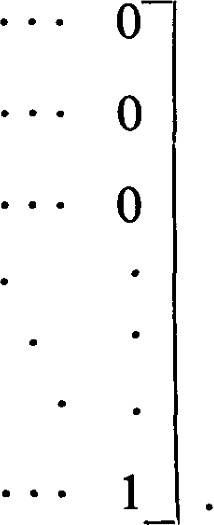
Type I: Multiply a row of A by a non-zero scalar.

Type Il: Add a multiple of one row to another row. Type Ill: Interchange two rows.

Elementary column operations are defined in an analogous way.

From a logical point of view these operations are redundant. An operation of type Ill can be accomplished by a combination of operations of types I and Il. It would, however, require four such operations to take the place of one operation of type Ill. Since we wish to develop convenient computational methods, it would not suit our purpose to reduce the number of operations at our disposal. On the other hand, it would not be of much help to extend the list of operations at this point. The student will find that, with practice, he can combine several elementary operations into one step. For example, such a combined operation would be the replacing of a row by a linear combination of rows, provided that the row replaced appeared in the linear combination with a non-zero coeffcient. We leave such short cuts to the student.

An elementary operation can also be accomplished by multiplying A on the left by a matrix. Thus, for example, multiplying the second row by the scalar c can be effected by the matrix

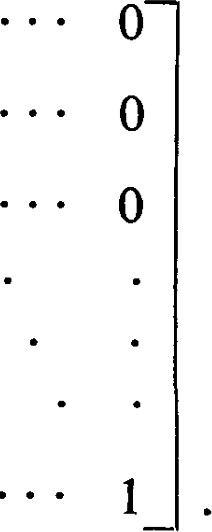
-1 O 0 c 0

* + 1

E2(c) =(6.1)

* + o

The addition of k times the third row to the first row can be effected by the matrix

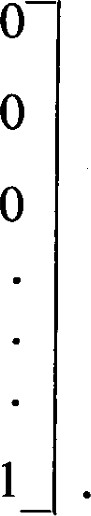
1 0 k o 1 0

## o o 1

E31(k) =(6.2)

O 0 0

The interchange of the first and second rows can be effected by the matrix

1 0 I O 0



-o

0

0 o 1

12 —(6.3)

### 0 0

These matrices corresponding to the elementary operations are called elementary matrices. These matrices are all non-singular and their inverses

are also elementary matrices. For example, the inverses of E2(c), E31(k), and E12 are respectively  and E12.

Notice that the elementary matrix representing an elementary operation is the matrix obtained by applying the elementary operation to the unit matrix.

Theorem 6.1. Any non-singular matrix A can be written as a product of elementary matrices.

PROOF. At least one element in the first column is non-zero or else A would be singular. Our first goal is to apply elementary operations, if necessary, to obtain a 1 in the upper left-hand corner. If all = 0, we can interchange rows to bring a non-zero element into that position. Thus we may as well suppose that an 0. We can then multiply the first row by an—I . Thus, to simplify notation, we may as well assume that all We now add —ail times the first row to the ith row to make every other element in the first column equal to zero.

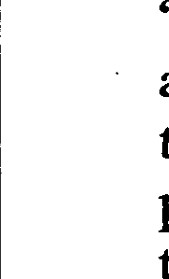
The resulting matrix is still non-singular since the elementary operations applied were non-singular. We now wish to obtain a 1 in the position of element a22. At least one element in the second column other than a12 is non-zero for otherwise the first two columns would be dependent. Thus by a possible interchange of rows, not including row 1, and multiplying the second row by a non-zero scalar we can obtain a22 = 1. We now add —an times the second row to the ith row to make every other element in the second column equal to zero. Notice that we also obtain a O in the position of a12 without affecting the 1 in the upper left-hand corner.

We continue in this way until we obtain the identity matrix. Thus if Er are elementary matrices representing the successive elementary operations, we have

1 = • E2EIA,

or(6.4) A = El-IE2 E-1.

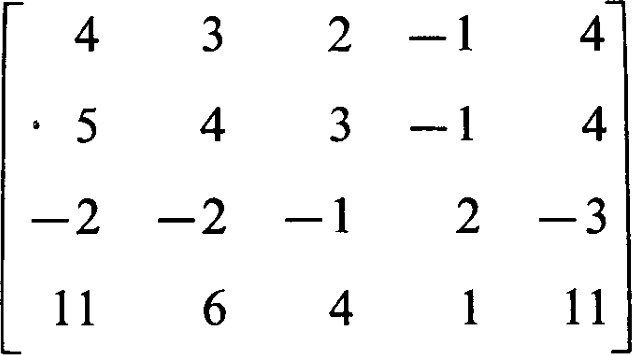
In Theorem 5.1 we obtained the Hermite normal form A' from the matrix A by multiplying on the left by the non-singular matrix Q—l . We see now that is a product of elementary matrices, and therefore that A can be transformed into Hermite normal form by a succession of elementary row operations. It is most emcient to use the elementary row operations directly without obtaining the matrix Q--1

We could have shown directly that a matrix could be transformed into Hermite normal form by means of elementary row operations. We would then be faced with the necessity of showing that the Hermite normal form obtained is unique and not dependent on the particular sequence of operations used. While this is not particularly diffcult, the demonstration is uninteresting and unilluminating and so tedious that it is usually left as an 'exercise for the reader." Uniqueness, however, is a part of Theorem 5.1, and we are assured that the Hermite normal form will be independent of the particular sequence of operations chosen. This is important as many possible operations are available at each step of the work, and we are free to choose those that are most convenient.

Basically, the instructions for reducing a matrix to Hermite normal form are contained in the proof of Theorem 6.1. In that theorem, however, we were dealing with a non-singular matrix and thus assured that we could at certain steps obtain a non-zero element on the main diagonal. For a singular matrix, this is not the case. When a non-zero element cannot be obtained with the instructions given we must move our consideration to the next column.

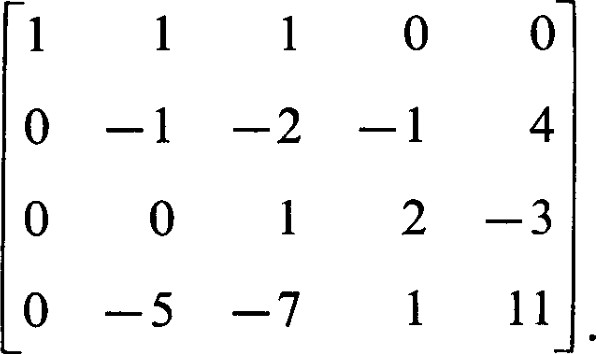
In the following example we perform several operations at each step to conserve space. When several operations are performed at once, some care must be exercised to avoid reducing the rank. This may occur, for example, if we subtract a row from itself in some hidden fashion. In this example we avoid this pitfall, which can occur when several operations of type Ill are combined, by considering one row as an operator row and adding multiples of it to several others.

Consider the matrix

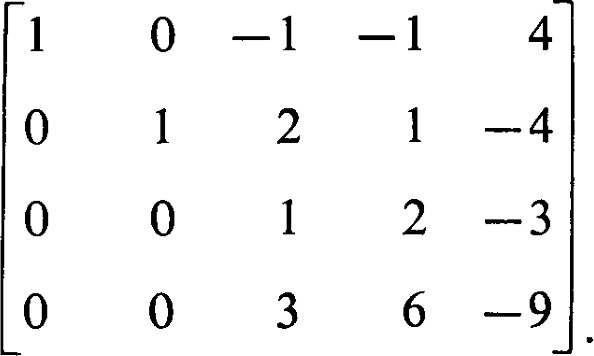


as an example.

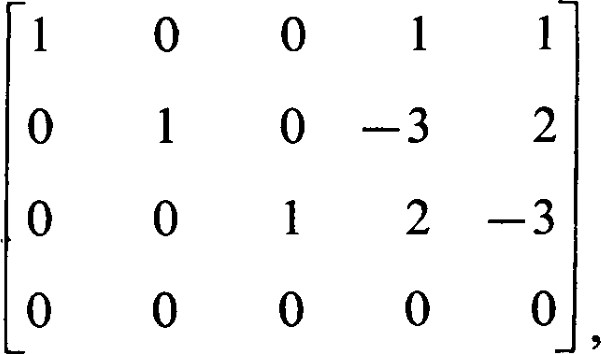
According to the instructions for performing the elementary row operations we should multiply the first row by å. To illustrate another possible way to obtain the "l" in the upper left corner, multiply row 1 by —1 and add row 2 to row l. Multiples of row I can now be added to the other rows to obtain



Now, multiply row 2 by —1 and add appropriate multiples to the other rows to obtain

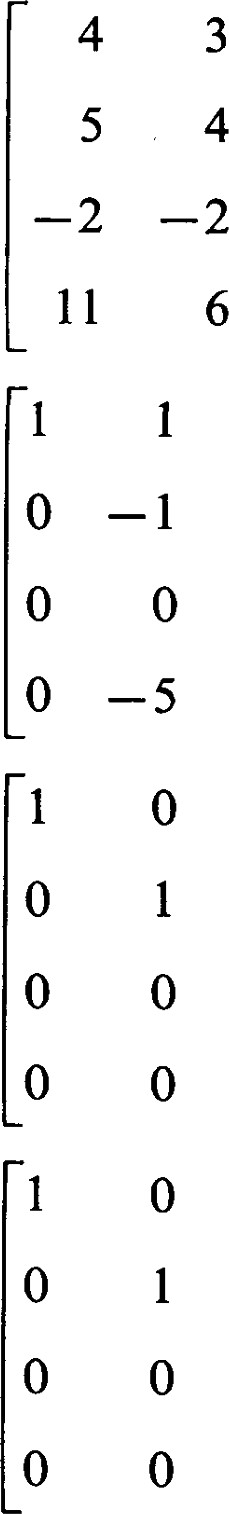
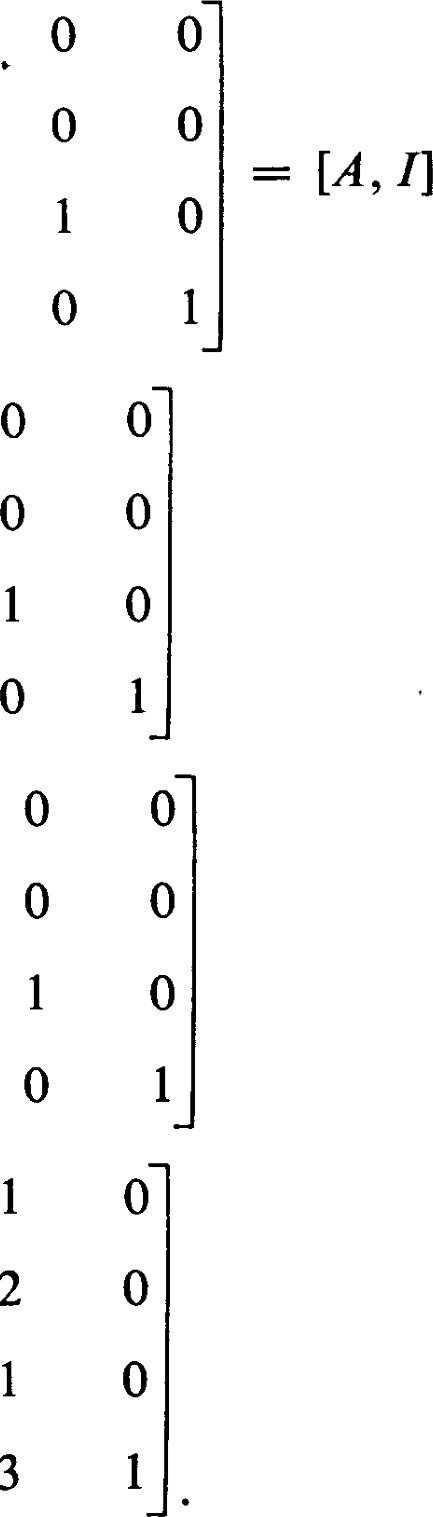


Finally, we obtain



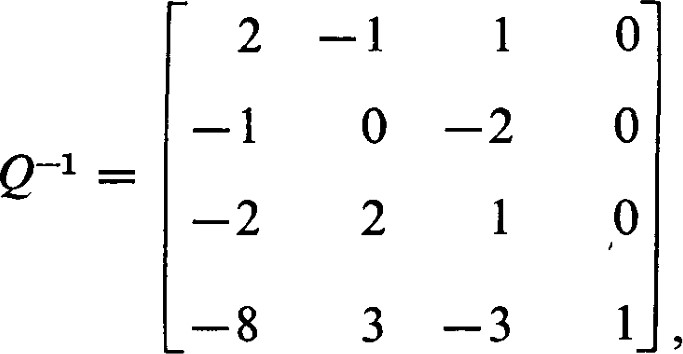
which is the Hermite normal form described in Theorem 5.1. If desired, can be obtained by applying the same sequence of elementary row operations to the unit matrix. However, while the Hermite normal form is necessarily unique, the matrix need not be unique, as the proof of Theorem 5.1 should show.

Rather than trying to remember the sequence of elementary operations used to reduce A to Hermite normal form, it is more emcient to perform these operations on the unit matrix at the same time we are operating on A. It is suggested that we arrange the work in the following way:

  11 11 —11

9 14 

In the end we obtain

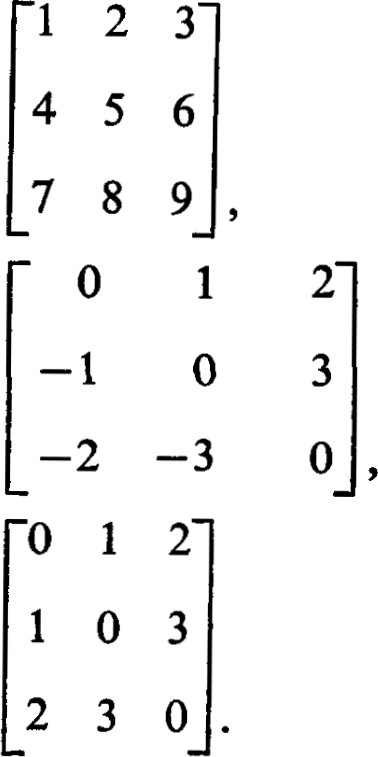


Verify directly that Q—IA is in Hermite normal form.

If A were non-singular, the Hermite normal form obtained would be the identity matrix. In this case would be the inverse of A. This method of finding the inverse of a matrix is one of the easiest available for hand computation. It is the recommended technique.

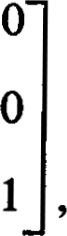
EXERCISES

I. Elementary operations provide the easiest methods for determining the rank of a matrix. Proceed as if reducing to Hermite normal form. Actually, it is not necessary to carry out all the steps as the rank is usually evident long before the

Hermite normal form is obtained. Find the ranks of the following matrices: (a)

(b)

### (c)

1. Identify the elementary operations represented by the following elementary matrices :
   1. 1 O



o

O 1

—2 o

* 1. o 1

1 O o o ,

* 1. o

2

O o

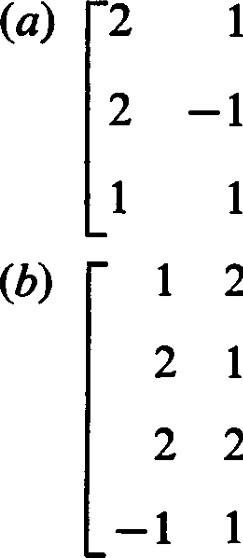
1. Show that the product

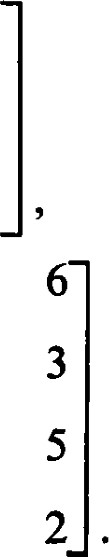
—1 0 1 0 1 —1

# 1 1 0 1 1 1

is an elementary matrix. Identify the elementary operations represented by each matrix in the product.

1. Show by an example that the product of elementary matrices is not necessarily an elementary matrix.
2. Reduce each of the following matrices to Hermite normal form.

3 -2



2

5

5

5 2

1 1 3 3 10 6 o o 2 1

3 2

1. Use elementary row operations to obtain the inverses of

(a)-1

2 , and

## (b) 1 2 3

7. (a) Show that, by using a sequence of elementary operations of type Il only, any two rows of a matrix can be interchanged with one of the two rows multiplied by —1. (In fact, the type Il operations involve no scalars other than ± 1.)

1. Using the results of part (a), show that a type Ill operation can be obtained by a sequence of type Il operations and a single type I operation.
2. Show that the sign of any row can be changed by a sequence of type Il operations and a single type Ill operation.

8. Show that any matrix A can be reduced to the form described in Theorem 4.1 by a sequence of elementary row operations and a sequence of elementary column operations.

7 1 Linear Problems and Linear Equations

For a given linear transformation of U into V and a given e V the problem of finding any or all e U for which = is called a linear problem. Before providing any specific methods for solving such problems, let us see what the set of solutions should look like.

If c(U), then the problem has no solution.

If e c(U), the problem has at least one solution. Let be one such solution. We call any such a particular solution. If is any other solution, then — geo) = c(é) — c(éo) = — = O so that — is in the kernel Ofc. Conversely, if E — his in the kernel of c then Ü(E) = c(éo + — $0) — c(éo) + — geo) = + O = so that is a solution. Thus the set of all solutions of c(é) = is of the form

 (7.1)

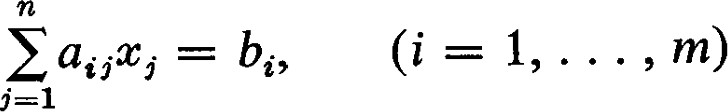
Since {$0} contains just one element, there is a one-to-one correspondence between the elements of K(c) and the elements of {h} + K(c). Thus the size of the set of solutions can be described by giving the dimension of K(c). The set of all solutions of the problem Ü(E) = is not a subspace of U unless

 = 0. Nevertheless, it is convenient to say that the set is of dimension v, the nullity of c.

Given the linear problem 0($) = F, the problem = 0 is called the associated homogeneous problem. The general solution is then any particular solution plus the solution of the associated homogeneous problem. The solution of the associated homogeneous problem is the kernel of c.

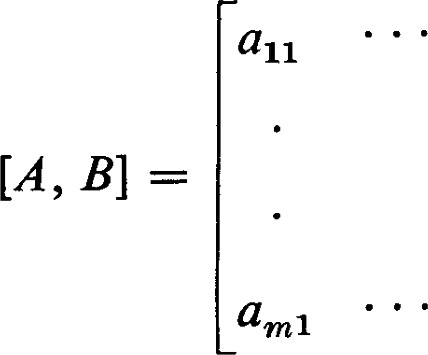
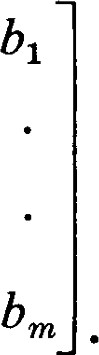
Now let be represented by the m x n matrix A = [as.j], be represented by B = (bl, , bm), and by X = ($1, x ). Then the linear problem c(é) = 18 becomes

AX = B (7.2) in matrix form, or

(7.3)

in the form of a system of linear equations.

Given A and B, the augmented matrix [A, B] of the system of linear equations is defined to be

aln

(7.4)

amn

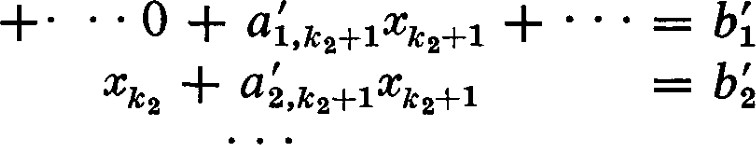
Theorem 7.1. The system of simultaneous linear equations AX = B has a solution if and only if the rank of A is equal to the rank of the augmented matrix [A, B]. Whenever a solution exists, all solutions can be expressed in terms of v = n — p independent parameters, where p is the rank of A.

PROOF. We have already seen that the linear problem = has a solution if and only if e c(U). This is the case if and only if is linearly dependent on {c(oq), c(æn)}. But this is equivalent to the condition that B be linearly dependent on the columns of A. Thus adjoining the column of bi's to form the augmented matrix must not increase the rank. Since the rank of the augmented matrix cannot be less than the rank of A we see that the system has a solution if and only if these two ranks are equal.

Now let Q be a non-singular matrix such that Q—IA = A' is in Hermite normal form. Any solution of AX = B is also a solution of A'X = Q—IAX — 0—1B = B'. Conversely, any solution of A'X = B' is also a solution of

AX = QA'X = QB' = B. Thus the two systems of equations are equivalent.

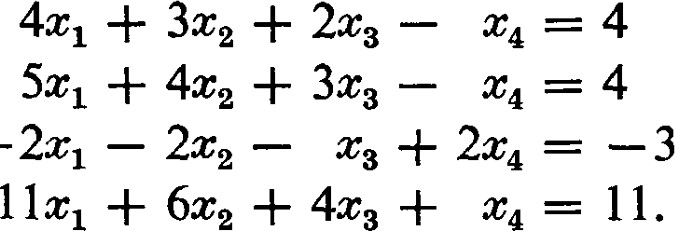
Now the system A' X = B' is particularly easy to solve since the variable appears only in the ith equation. Furthermore, non-zero coeffcients appear only in the first p equations. The condition that e c(U) also takes on a form that is easily recognizable. The condition that B' be expressible as a linear combination of the columns of A' is simply that the elements of B' below row p be zero. The system A' X = B' has the form

 + al, k1+1 k1+1

(7.5)

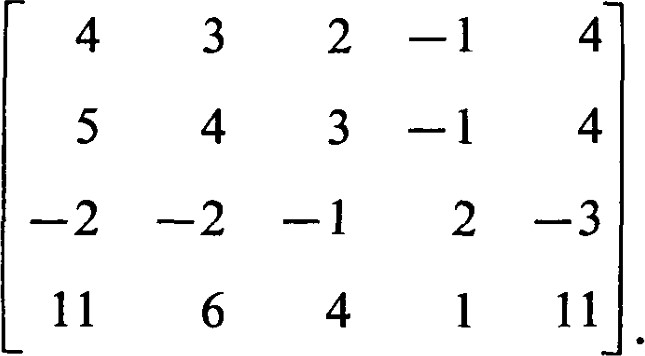
Since each appears in but one equation with unit coefficient, the remaining n — p unknowns can be given values arbitrarily and the corresponding values of the computed. The n — p unknowns with indices not the ki are the n — p parameters mentioned in the theorem. 

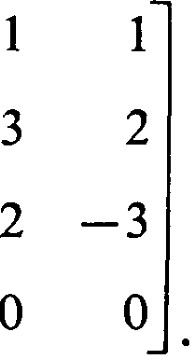
As an example, consider the system of equations:



—

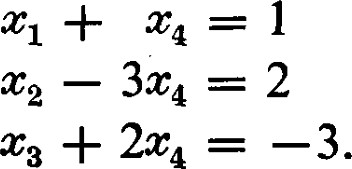
The augmented matrix is



This is the matrix we chose for an example in the previous section. There we obtained the Hermite normal form 1 0 o 0 1 0 0 1

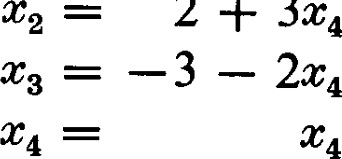
### 0 o 0

Thus the system of equations A'X = B' corresponding to this augmented matrix is



It is clear that this system is very easy to solve. We can take any value whatever for and compute the corresponding values for $1, $2, and $3. A particular solution, obtained by taking 

It is more instructive to write the new system of equations in the form



2

+

3X4

In vector form this becomes



We can easily verify that ( l , 3, —2, l) is a solution of the associated homogeneous problem. In fact, {(—1, 3, —2, l)} is a basis for the kernel, and 3, —2, 1), for an arbitrary $4, is a general element of the kernel. We have, therefore, expressed the general solution as a particular solution plus the kernel.

The elementary row operations provide us with the recommended technique for solving simultaneous linear equations by hand. This application is the principal reason for introducing elementary row operations rather than column operations.

Theorem 7.2. The equation AX = B fails to have a solution if and only if there exists a one-row matrix C such that CA = 0 and CB = 1.

PROOF. Suppose the equation AX = B has a solution and a C exists such that CA = O and CB = 1. Then we would have 0 = (CA)X = C(AX) = CB = l , which is a contradiction.

On the other hand, suppose the equation AX = B has no solution. By Theorem 7.1 this implies that the rank of the augmented matrix [A, B] is greater than the rank of A. Let Q be a non-singular matrix such that Q-I [A, B] is in Hermite normal form. Then if p is the rank of A, the (p + l)st row of Q-I [A, B] must be all zeros except for a I in the last column. If C is the (p + l)st row of this means that

CIA, B] = [0 0 or and CD = 1. 

This theorem is important because it provides a positive condition for a negative conclusion. Theorem 7.1 also provides such a positive condition and it is to be preferred when dealing with a particular system of equations. But Theorem 7.2 provides a more convenient condition when dealing with systems of equations in general.

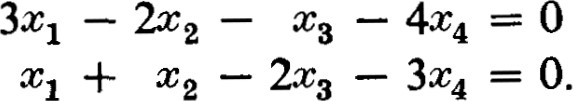
Although the sytems of linear equations in the exercises that follow are written in expanded form, they are equivalent in form to the matric equation AX = B. From any linear problem in this set, or those that will occur later, it is possible to obtain an extensive list of closely related linear problems that appear to be different. For example, if AX = B is the given linear problem with A an m x n matrix and Q is any non-singular m x m matrix, then A'X = B' with A' = QA and B' = QB is a problem with the same set of solutions. If P is a non-singular n x n matrix, then A"X" = B where A" = AP is a problem whose solution X" is related to the solution X of the original problem by the condition X" = P—IX.

For the purpose of constructing related exercises of the type mentioned, it is des-irable to use matrices P and Q that do not introduce tedious numerical calculations. It is very easy to obtain a non-singular matrix P that has only integral elements and such that its inverse also has only integral elements. Start with an identity matrix of the desired order and perform a sequence of elementary operations of types Il and Ill. As long as an operation of type I is avoided, no fractions will be introduced. Furthermore, the inverse operations will be of types Il and Ill so the inverse matrix will also have only integral elements.

For convenience, some matrices with integral elements and inverses with integral elements are listed in an appendix. For some of the exercises that are given later in this book, matrices of transition that satisfy special conditions are also needed. These matrices, known as orthogonal and unitary matrices, usually do not have integral elements. Simple matrices of these types are somewhat harder to obtain. Some matrices of these types are also listed in the appendix.

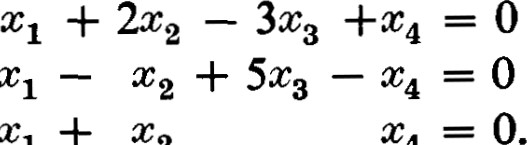
EXERCISES

1. Show that {(1, 1, 1, 0), (2, 1, 0, 1)} spans the subspace of all solutions of the system of linear equations



1. Find the subspace of all solutions of the system of linear equations

4



3X1

2X1

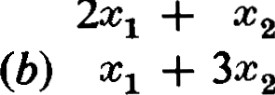
+

1. Find all solutions of the following two systems of non-homogeneous linear equations.

 + + — 234 — 11

3X1 — 2X2 — '7X3 + 5X4 = 

+



2$1

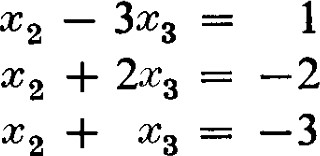
3X1 — 2.C2 — 2X1

+ 5X4 = 10

+ 4x

+ 5x4 = 5.

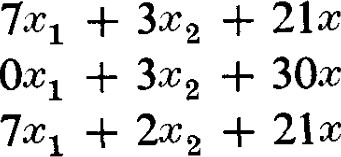
1. Find all solutions of the following system of non-homogeneous linear equations

4X1 — 3x2 

1 3.

1. Find all solutions of the system of equations,

### 3 — 13X4 5 = — 14



IOXI

— 16T4 + = —23

3 — llX4 = —16

9X1 + 3C2 + 27x3 — 15x4 + x = —20.

1. Theorem 7.1 states that a necessary and suffcient condition for the existence of a solution of a system of simultaneous linear equations is that the rank of the augmented matrix be equal to the rank of the coemcient matrix. The most effcient way to determine the rank of each of these matrices is to reduce each to Hermite normal form. The reduction of the augmented matrix to normal form, however, automatically produces the reduced form of the coemcient matrix. How, and where? How is the comparison of the ranks of the coemcient matrix and the augmented matrix evident from the appearance of the reduced form of the augmented matrix ?
2. The differential equation d 2y/dx2 + 4y = sin x has the general solution y = Cl sin + C2 cos 2x + sin x. Identify the associated homogeneous problem, the solution of the associated homogeneous problem, and the particular solution.

8 1 Other Applications of the Hermite Normal Form

The Hermite normal form and the elementary row operations provide techniques for dealing with problems we have already encountered and handled rather awkwardly.

#### A Standard Basisfor a Subspace

Let A } be a basis of U and let W be a subspace of U spanned by the set B ß }. Since every subspace of U is spanned by a finite set, it is no restriction to assume that B is finite. Let = E so that (b bin) is the n-tuple representing L. Then in the matrix B = [bt,.] each row is the representation of a vector in B. Now suppose an elementary row operation is applied to B to obtain B'. Every row of B' is a linear combination of the rows of B and, since an elementary row operation has an inverse, every row of B is a linear combination of the rows of B'. Thus the rows of B and the rows of B' represent sets spanning the same subspace W. We can therefore reduce B to Hermite normal form and obtain a particular set spanning W. Since the non-zero rows of the Hermite normal form are linearly independent, they form a basis of W.

Now let C be another set spanning W. In a similar fashion we can construct a matrix C whose rows represent the vectors in C and reduce this matrix to Hermite normal form. Let C' be the Hermite normal form obtained from C, and let B' be the Hermite normal form obtained from B. We do not assume that B and C have the same number of elements, and therefore B' and C' do not necessarily have the same number of rows. However, in each the number of non-zero rows must be equal to the dimension of W. We claim that the non-zero rows in these two normal forms are identical.

To see this, construct a new matrix with the non-zero rows of C' written beneath the non-zero rows of B' and reduce this matrix to Hermite normal form. Since the rows of C' are dependent on the rows of B', the rows of C' can be removed by elementary operations, leaving the rows of B'. Further reduction is not possible since B' is already in normal form. But by interchanging rows, which are elementary operations, we can obtain a matrix in which the non-zero rows of B' are beneath the non-zero rows of C'. As before, we can remove the rows of B' leaving the non-zero rows of C' as the normal form. Since the Hermite normal form is unique, we see that the non-zero rows of B' and C' are identical. The basis that we obtain from the non-zero rows of the Hermite normal form is the standard basis with respect to A for the subspace W.

This gives us an effective method for deciding when two sets span the same subspace. For example, in Chapter 1-4, Exercise 5, we were asked to show that {(1, 1, 0, 0), (l, 0, l , 1)} and {(21, —1)} span the same space. In either case we obtain {(1, 0, l , l), (0, l , —1, as the standard basis.

The Sum of Two Subspaces

If Al is a subset spanning WI and A2 is a subset spanning W2, then Al U A2 spans WI + W2 (Chapter I, Proposition 4.4). Thus we can find a basis for WI + W2 by constructing a large matrix whose rows are the representations of the vectors in Al U .A2 and reducing it to Hermite normal form by elementary row operations.

The Characterization of a Subspace by a Set of Homogeneous Linear Equations

We have already seen that the set of all solutions of a system of homogeneous linear equations is a subspace, the kernel of the linear transformation represented by the matrix of coemcients. The method for solving such a system which we described in Section 7 amounts to passing from a characterization of a subspace as the set of all solutions of a system of equations to its description as the set of all linear combinations of a basis. The question naturally arises: If we are given a spanning set for a subspace W, how can we find a system of simultaneous homogeneous linear equations for which W is exactly the set of solutions ?

This is not at all diffcult and no new procedures are required. All that is needed is a new look at what we have already done. Consider the homogeneous linear equation alX1 -F • • • + ancn = 0. There is no significant difference between the a/s and the Xi's in this equation; they appear symmetrically. Let us exploit this symmetry systematically.

= O and + • • • + bnxn — \_ O are two homogeneous linear equations then (al + bl)X1 + • • • + (an + bn)xn = 0 is a homogeneous linear equation as also is aaltl + • • • + aanxn —— 0 where a e F. Thus we can consider the set of all homogeneous linear equations in n unknowns as a vector space over F. The equation alti + • • • + anxn is represented by the n-tuple (a , an).

When we write a matrix to represent a system of equations and reduce that matrix to Hermite normal form we are finding a standard basis for the subspace of the vector space of all homogeneous linear equations in x spanned by this system of equations just as we did in the first part of this section for a set of vectors spanning a subspace. The rank of the system of equations is the dimension of the subspace of equations spanned by the given system.

Now let W be a subspace given by a spanning set and solve for the subspace E of all equations satisfied by W. Then solve for the subspace of solutions of the system of equations E. W must be a subspace of the set of all solutions. Let W be of dimension v. By Theorem 7.1 the dimension of E is n — v. Then, in turn, the dimension of the set of all solutions of E isn — (n — v) Thus W must be exactly the space of all solutions. Thus W and E characterize each other.

If we start with a system of equations and solve it by means of the Hermite normal form, as described in Section 7, we obtain in a natural way a basis for the subspace of solutions. This basis, however, will not be the standard basis. We can obtain full symmetry between the standard system of equations and the standard basis by changing the definition of the standard basis. Instead of applying the elementary row operations by starting with the lefthand column, start with the right-hand column. If the basis obtained in this way is called the standard basis, the equations obtained will be the standard equations, and the solution of the standard equations will be the standard basis. In the following example the computations will be carried out in this way to illustrate this idea. It is not recommended, however, that this be generally done since accuracy with one definite routine is more important.

Let w = o, —3, 11, —5), (3, 2, 5, —5, 3), (1, 1, 2, —4, 2), (7, 2, 12, 1, We now find a standard basis by reducing

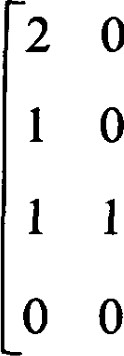
1I l —5

3 2  3

12

7 2 12 1 2

to the form

5 o

2 1

#### O o o

O o o

From this we see that the coemcients of our systems of equations satisfy the

|  |  |  |
| --- | --- | --- |
| conditions |  |  |
| 2a1 | + 5aa | 0 |

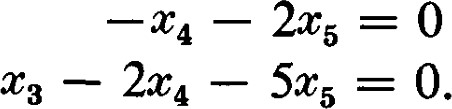
al + 2a3 -+ a4 

##### al +

The coemcients al and a3 can be selected arbitrarily and the others computed from them. In particular, we have

(al, a2, a3, a4, %) = a (l l, —2) a3(0 0 1

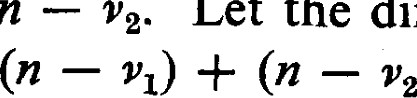
The 5-tuples (l 1 0 1, —2) and (0, O I , —2, —5) represent the two standard linear equations



The reader should check that the vectors in W actually satisfy these equations and that the standard basis for W is obtained.

The Intersection of Two Subspaces

Let WI and W2 be subspaces of U of dimensions VI and "2, respectively, and let WI n W2 be of dimension v. Then WI + W2 is of dimension VI + 122 — v. Let El and E2 be the spaces of equations characterizing WI and W2. As we have seen El is of dimension n — and E2 is of dimension dimension of El + F-2 be p. Then El n E2 is of dimension ) — p = 2n — PI —



n

—

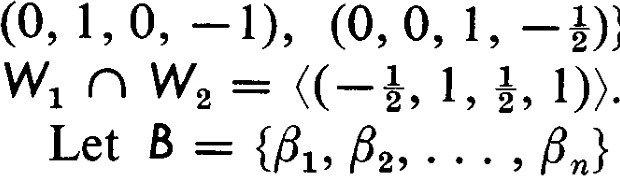
Let

the

Since the vectors in WI n W2 satisfy the equations in both El and E2, they satisfy the equations in El + E2. Thus v n — p. On the other hand, WI and W2 both satisfy the equations in El n E2 so that WI + W2 satisfies the equations in El n E2. Thus VI + — v n — {2n 

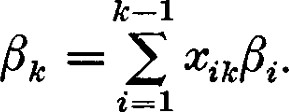
+ + p — n. A comparison of these two inequalities shows that v = n — p and hence that WI n W2 is characterized by El + E2.

Given WI and W2, the easiest way to find WI n W2 is to determine El and E2 and then El + E2. From El + E2 we can then find WI n W2. In effect, this involves solving three systems of equations, and reducing to Hermit normal form three times, but it is still easier than a direct assault on the problem.

As an example consider Exercise 8 of Chapter 1-4. Let WI = 2, 3, 6), (4, —1, 3, 6), (5, l, 6, 12)) and —l, 1, l), (2, —1, 4, Using the Hermite normal form, we find that El and E2 — — —3, 0, l), (—3, —2, 1, O)). Again, using the Hermite normal form we find that the standard basis for El -F E2 is {(1, 0, 0, å), —I)}. And from this we find quite easily that,

} be a given finite set of vectors. We wish to solve the problem posed in Theorem 2.2 of Chapter I. How do we show that some is a linear combination of the with i < k; or how do we show that no can be so represented ?

We are looking for a relation of the form

 (8.1)

This is not a meaningful numerical problem unless is a given specific set. This usually means that the are given in terms of some coordinate system, relative to some given basis. But the relation (8.1) is independent of any coordinate system so we are free to choose a different coordinate system if this will make the soluti'on any easier. It turns out that the tools to solve this problem are available.

Let A = {} be the given basis and let

ij i'  n. (8.2)

If A' — , u; } is the new basis (which we have not specified yet), we would have

ij i'  n. (8.3)

What is the relation between A = [at,] and A' = If P is the matrix of transition from the basis A to the basis A', by formula (4.3) we see that

A = PA'. (8.4)

Since P is non-singular, it can be represented as a product of elementary matrices. This means A' can be obtained from A by a sequence of elementary row operations.

The solution to (8.1) is now most conveniently obtained if we take A' to be in Hermite normal form. Suppose that A' is in Hermite normal form and use the notation given in Theorem 5.1. Then, for we would have

 — oc'i, (8.5) and for j between kr and kT+1 we would have

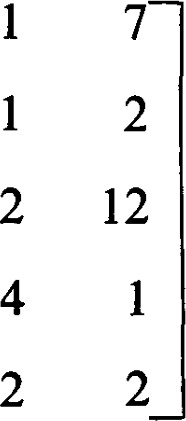
aijoti

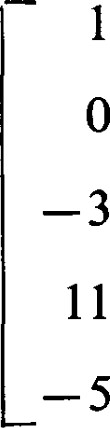
= a;jßki (8.6)

Since ki < kr < j, this last expression is a relation of the required form. (Actually, every linear relation that exists among the can be obtained from those in (8.6). This assertion will not be used later in the book so we will not take space to prove it. Consider it "an exercise for the reader." ) Since the columns of A and A' represent the vectors in B, the rank of A is equal to the number of vectors in a maximal linearly independent subset of B. Thus, if B is linearly independent the rank of A will be n, this means that the Hermite normal form of A will either show that B is linearly independent or reveal a linear relation in B if it is dependent.

For example, consider the set {(1, O, —3 Il

(l, l , 2, —4, 2), (7, 2, 12, l, 2)}. The implied context is that a basis A = } is considered to be given and that — + Il 014 — etc. According to (8.2) the appropriate matrix is

3



r

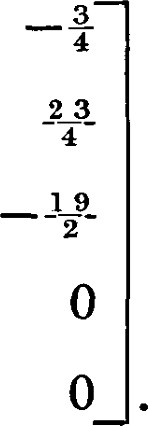
2

5

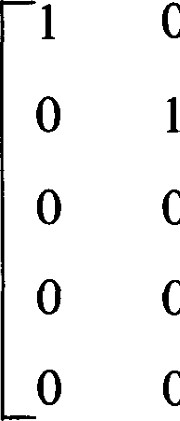
3

which reduces to the Hermite normal form

0



3



0

0

o

0

0

1

O

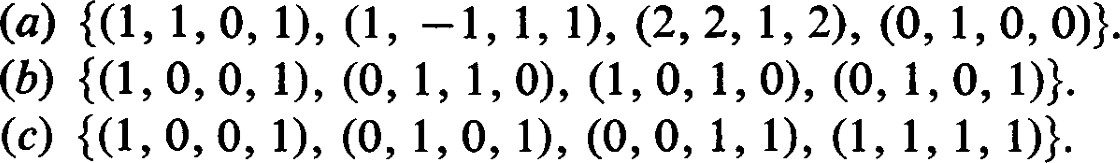
0

It is easily checked that

##### -1-9-(1, 1, 2, —4, 2) = (7, 2, 12, 2)

EXERCISES

1. Determine which of the following set in R4 are linearly independent over R.



This problem is identical to Exercise 8, Chapter 1-2.

1. Let W be the subspace of R5 spanned by {(1, 1, 1, 1, 1), (1, O, 1, O, 1),



Find a standard basis for W and the dimension of W. This problem is identical to Exercise 6, Chapter 1-4.

1. Show that {(1, 1, 2, —3), (1, 1, 2, 0), (3, —1, 6, and {(1, 0, 1, 0), (O, 2, O, 3)} do not span the same subspace. This problem is identical to Exercise 7, Chapter 1-4.
2. If WI = 1, 3, -1), (1, O, —2, 0), (3, 2, 4, and  O, O, 1), (1, 1, 7, 1)) determine the dimension of WI + W2.
3. Let w = 1, —3, o, 1), (2, 1, o, —1, 4), (3, 1, -1, 1, 8), (1, 2, 3, 2, Determine the standard basis for W. Find a set of linear equations which characterize W.
4. Let WI = 2, 3, 6), (4, —1, 3, 6), (5, 1, 6, 12)) and = —1, 1, 1), (2, —1, 4, 5)) be subspaces of R4. Find bases for WI n W2 and WI + W2. Extend the basis of WI n W2 to a basis of WI and extend the basis of WI n W2 to a basis of W2. From these bases obtain a basis of WI + W2. This problem is identical to Exercise 8, Chapter 1-4.

9 1 Normal Forms

To understand fully what a normal form is, we must first introduce the concept of an equivalence relation. We say that a relation is defined in a set if, for each pair (a, b) of elements in this set, it is decided that "a is related to b" or "a is not related to b." If a is related to b, we write a rw b. An equivalence relation in a set S is a relation in S satisfying the following laws :

Reflexive law: a a,

Symmetric law: If a b, then b a.

Transitive law: If a b and b c, then a c.

If for an equivalence relation we have a b, we say that a is equivalent to b.

9 Normal Forms

Examples. Among rational fractions we can define a/b c/d ( for a, b, c, d integers) if and only if ad = bc. This is the ordinary definition of equality in rational numbers, and this relation satisfies the three conditions of an equivalence relation.

In geometry we do not ordinarily say that a straight line is parallel to itself. But if we agree to say that a straight line is parallel to itself, the concept of parallelism is an equivalence relation among the straight lines in the plane or in space.

Geometry has many equivalence relations: congruence of triangles, similarity of triangles, the concept of projectivity in projective geometry, etc. In dealing with time we use many equivalence relations: same hour of the day, same day of the week, etc. An equivalence relation is like a generalized equality. Elements which are equivalent share some common or underlying property. As an example of this idea, consider a collection of sets. We say that two sets are equivalent if their elements can be put into a one-to-one correspondence; for example, a set of three battleships and a set of three cigars are equivalent. Any set of three objects shares with any other set of three objects a concept which we have abstracted and called "three." All other qualities which these sets may have are ignored.

It is most natural, therefore, to group mutually equivalent elements together into classes which we call equivalence classes. Let us be specific about how this is done. For each a e S, let Sa be the set of all elements in S equivalent to a; that is, b e Sa if and only if b a. We wish to show that the various sets we have thus defined are either disjoint or identical.

Suppose Sa n .Sb is not empty; that is, there exists a c e Sa n Sb such that c —v a and b. By symmetry b c, and by transitivity b a. If d is any element of Sb, d "W b and hence a. Thus de Sa and Sb c S . Since the relation between Sa and Sb is symmetric we also have Sa c Sb and hence Sa = Sb. Since a e Sa we have shown, in effect, that a proposed equivalence class can be identified by any element in it. An element selected from an equivalence class will be called a representative of that class.

An equivalence relation in a set S defines a partition of that set into equivalence classes in the following sense: (l) Every element of S is in some equivalence class, namely, a e So. (2) Two elements are in the same equivalence class if and only if they are equivalent. (3) Non-identical equivalence classes are disjoint. On the other hand, a partition of a set into disjoint subsets can be used to define an equivalence relation; two elements are equivalent if and only if they are in the same subset.

The notions of equivalence relations and equivalence classes are not nearly so novel as they may seem at first. Most students have encountered these ideas before, although sometimes in hidden forms. For example, we may say that two differentiable functions are equivalent if and only if they have the same derivative. In calculus we use the letter "C" in describing the equivalence classes; for example, \*3 + x2 + 2x + C is the set (equivalence class) of all functions whose derivative is 3+ + 2x + 2.

In our study of matrices we have so far encountered four different equivæ lence relations :

I. The matrices A and B are said to be left associate if there exists a non-singular matrix Q such that B = Q—IA. Multiplication by corresponds to performing a sequence of elementary row operations. If A represents a linear transformation c of U into V with respect to a basis A in U and a basis B in V, the matrix B represents with respect to A and a new basis in V.

Il. The matrices A and B are said to be right associate if there exists a non-singular matrix P such that B = AP.

Ill. The matrices A and B are said to be associate if there exist nonsingular matrixes P and Q such that B = Q—IAP. The term "associate" is not a standard term for this equivalence relation, the term most frequently used being "equivalent." It seems unnecessarily confusing to use the same term for one particular relation and for a whole class of relations. Moreover, this equivalence relation is perhaps the least interesting of the equivalence relations we shall study.

IV. The matrices A and B are said to be similar if there exists a nonsingular matrix P such that B = P —IAP. As we have seen (Section 4) similar matrices are representations of a single linear transformation of a vector space into itself. This is one of the most interesting of the equivalence relations, and Chapter Ill is devoted to a study of it.

Let us show in detail that the reation we have defined as left associate is an equivalence relation. The matrix appears in the definition because Q represents the matrix of transition. However, Q 1 is just another singular matrix, so it is clearly the same thing to say that A and B are left associate if and only if there exists a non-singular matrix Q such that B = QA. (l) ANA since IA = A.

1. If A B, there is a non-singular matrix Q such that B = QA. But then A = Q—IB so that B A.
2. If A 'v B and B N C, there exist non-singular matrices Q and P such that B = QA and C = PB. But then PQA = PB = C and PQ is non-singular so that A C.

For a given type of equivalence relation among matrices a normal form is a particular matrix chosen from each equivalence class. It is a representative of the entire class of equivalent matrices. In mathematics the terms "normal" and "canonical" are frequently used to mean "standard' in some particular sense. A normal form or canonical form is a standard

9 Normal Forms

form selected to represent a class of equivalent elements. A normal form should be selected to have the following two properties: Given any matrix A, (l) it should be possible by fairly direct and convenient methods to find the normal form of the equivalence class containing A, and (2) the method should lead to a unique normal form.

Often the definition of a normal form is compromised with respect to the second of these desirable properties. For example, if the normal form were a matrix with complex numbers in the main diagonal and zeros elsewhere, to make the normal form unique it would be necessary to specify the order of the numbers in the main diagonal. But it is usually sumcient to know the numbers in the main diagonal without regard to their order, so it would be an awkward complication to have to specify their order.

Normal forms have several uses. Perhaps the most important use is that the normal form should yield important or useful information about the concept that the matrix represents. This should be amply illustrated in the case of the concept of left associate and the Hermite normal form. We introduced the Hermite normal form through linear transformations, but we found that it yielded very useful information when the matrix was used to represent linear equations or bases of subspaces.

Given two matrices, we can use the normal form to tell whether they are equivalent. It is often easier to reduce each to normal form and compare the normal forms than it is to transform one into the other. This is the case, for example, in the application described in the first part of Section 8.



Sometimes, knowing the general appearance of the normal form, we can find all the information we need without actually obtaining the normal form. This is the case for the equivalence relation we have called associate. The normal form for this equivalence relation is described in Theorem 4.1. ere is just one normal form for each possible value of the rank. The itamber of different equivalence classes is min {m, n} + 1. With this notion of equivalence the rank of a matrix is the only property of importance. Any two matrices of the same rank are associate. In practice we can find the rank without actually computing the normal form of Theorem 4.1. And knowing the rank we know the normal form.

We encounter several more equivalence relations among matrices. The type of equivalence introduced will depend entirely on the underlying concepts the matrices are used to represent. It is worth mentioning that for the equivalence relations we introduce there is no necessity to prove, as we did for an example above, that each is an equivalence relation. An underlying concept will be defined without reference to any coordinate system or choice of basis. The matrices representing this concept will transform according to certain rules when the basis is changed. Since a given basis can be retained the relation defined is reflexive. Since a basis changed can be changed back to the original basis, the relation defined is symmetric. A basis changed once and then changed again depends only on the final choice so that the relation is transitive.

For a fixed basis A in U and B in V two different linear transformations c and T of U into V are represented by different matrices. If it is possible, however, to choose bases A' in U and B' in V such that the matrix representing T with respect to A' and B' is the same as the matrix representing with respect to A and B, then it is certainly clear that c and T share important geometric properties.

For a fixed two matrices A and A' representing with respect to different bases are related by a matrix equation of the form A' = Q—IAP. Since A and A' represent the same linear transformation we feel that they should have some properties in common, those dependent upon c.

These two points of view are really slightly different views of the same kind of relationship. In the second case, we can consider A and A' as representing two linear transformations with respect to the same basis, instead of the same linear transformation with respect to different bases.

1 0

For example, in R2 the matrix represents a reflection about the Cl-axis and represents a reflection about the xraxis. When both

linear transformations are referred to the same coordinate system they are different. However, for the purpose of discussing properties independent of a coordinate system they are essentially alike. The study of equivalence relations is motivated by such considerations, and the study of normal forms is aimed at determining just what these common properties are that are shared by equivalent linear transformations or equivalent matrices.

To make these ideas precise, let c and T be linear transformations of V into itself. We say that and T are similar if there exist bases A and B of V such that the matrix representing with respect to A is the same as the matrix representing T with respect to B. If A and B are the matrices representing  and T with respect to A and P is the matrix of transition from A to B, then P—IBP is the matrix representing T with respect to B. Thus c and T are similar if P—IBP = A.

In a similar way we can define the concepts of left associate, right associate, and associate for linear transformations.

\*10 1 Quotient Sets, Quotient Spaces

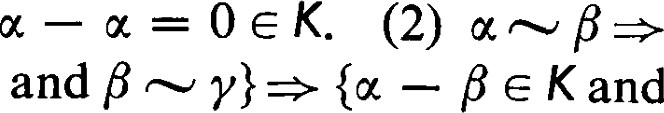
Definition. If S is any set on which an equivalence relation is defined, the collection of equivalence classes is called the quotient or factor set. Let S denote the quotient set. An element of S is an equivalence class. If a is an

10 Quotient Sets, Quotient Spaces

element of S and ä is the equivalence class containing a, the mapping that maps a onto ä is well defined. This mapping is called the canonical mapping.

Although the concept of a quotient set might appear new to some, it is certain that almost everyone has encountered the idea before, perhaps in one guise or another. One example occurs in arithmetic. In this setting, let S be the set of all formal fractions of the form a/b where a and b are integers and b 0. Two such fractions, a/b and c/d, are equivalent if and only if ad = bc. Each equivalence class corresponds to a single rational number. The rules of arithmetic provide methods of computing with rational numbers by performing appropriate operations with formal fractions selected from the corresponding equivalence classes.

Let U be a vector space over F and let K be a subspace of U. We shall call two vectors u, e U equivalent modulo K if and only if their difference lies in K. Thus if and only if oc — e K. We must first show this defines an equivalence relation. (l) because

— e K  (3) tN•' — e K and  — y e K}. Since K is a subspace — y = (u — P) + (F — y) e K and, hence, Y. Thus is an equivalence relation.

We wish to define vector addition and scalar multiplication in U. For e U, let e U denote the equivalence class containing u. is called a representative of a. Since may contain other elements besides u, it may happen that d and yet = a'. Let and be two elements in U. Since u, e U, + is defined. We wish to define + to be the sum of and B. In order for this to be well defined we must end up with the same equivalence class as the sum if different representatives are chosen from and B. Suppose = a' and B = B'. Then — Od G — K, and (oc

K. Thus -F = od -F F' and the sum is well defined. Scalar multiplication is defined similarly. For a e F, ai is thus defined to be the equivalence class containing am; that is, ai = acc These operations in U are said to be induced by the corresponding operation in U.

Theorem 10.1. If U is a vector space over F, and K is a subspace of U, the quotient set U with vector addition and scalar multiplication defined as above is a vector space over F.

PROOF. We leave this as an exercise. 

For any e U, the symbol oc + K is used to denote the set of all elements in U that can be written in the form oc + y where y e K. (Strictly speaking, we should denote the set by {u} + K so that the plus sign combines two objects of the same type. The notation introduced bere is traditional and simpler.) The set cc + K is called a coset of K. If + K, then — e K and

a. Conversely, if N 01, then — oc = G K so K. Thus  + K is simply the equivalence class containing 01. Thus + K = + K if and only if e B = + K or + K.

The notation + K to denote is convenient to use in some calculations. For example, + = (u + K) + (F + K) = + + K = + 13, and ai = a(oe, + K) = au + aK c au + K = acc. Notice that ai = au when and au are considered to be elements of U and scalar mutliplication is the induced operation, but that aa and au may not be the same when they are viewed as subsets of U (for example, let a = 0). However, since aä c acc the set ai determines the desired coset in U for the induced operations. Thus we can compute effectively in U by doing the corresponding operations with representatives. This is precisely what is done when we compute in residue classes of integers modulo an integer m.

Definition. U with the induced operations is called a factor space or quotient space. In order to designate the role of the subspace K which defines the equivalence relations, U is usually denoted by U/K.

In our discussion of solutions of linear problems we actually encountered quotient spaces, but the discussion was worded in such a way as to avoid introducing this more sophisticated concept. Given the linear transformation  of U into V, let K be the kernel of and let U = U/K be the corresponding quotient space. If and are solutions of the linear problem, oc($) = 18, then (T(otl — = 0 so that and are in the same coset of K. Thus for each 1B e Im(c) there corresponds precisely one coset of K. In fact the correspondence between U/K and Im(Ü) is an isomorphism, a fact which is made more precise in the following theorem.

Theorem 10.2. (First homomorphism theorem). Let be a linear transformation ofU into V. Let K be the kernel Ofc. Then can be written as the product of a canonical mapping of U onto U = U/K and a monomorphism of CJ into V.

PROOF. The canonical mapping has already been defined. To define 01, for each e U let (Tl(ä) = c(oc) where is any representative of a. Since c(u) = c(d) for d, is well defined. It is easily seen that is a monomorphism since must have different values in different cosets. 

The homomorphism theorem is usually stated by saying, "The homomorphic image is isomorphic to the quotient space of U modulo the kernel." Theorem 10.3. (Mapping decomposition theorem). Let c be a linear transformation ofU into V. Let K be the kernel of c and I the image Ofc. Then c can be written as the product c = tall'), where is the canonical mapping of

10 | Quotient Sets, Quotient Spaces 81

U onto U = IJ/K, is an isomorphism ofU onto l, and is the injection of I into V.

PROOF. Let d be the linear transformation of U onto I induced by restricting the codomain of c to the image of c. By Theorem 10.2, d can be written in the form d

Theorem 10.4. (Mapping factor theorem). Let S be a subspace of U and let U = U/S be the resulting quotient space. Let be a linear transformation of U into V, and let K be the kernel Ofc. If S c K, then there exists a linear transformation of U into V such that = where is the canonical mapping of U onto U.

PROOF. For each e U, let Cl(ä) = c(u) where e a. If od is another representative of a, then — u' e S c K. Thus o(oc) = o(u') and is well defined. It is easy to check that is linear. Clearly, c(u) = Ol(ä) = for all e U, and =

We say that c factors through U.

Note that the homomorphism theorem is a special case of the factor theorem in which K = S.

Theorem 10.5. (Induced mapping theorem). Let U and V be vector spaces over F, and let T be a linear transformation of U into V. Let (Jo be a subspace of U and let Vo be a subspace of V. If T(Uo) c Vo, it is possible to define in a natural way a mapping of U/Uo into V/ Vo such that = where is the canonical mapping U onto U and is the canonical mapping of V onto V.

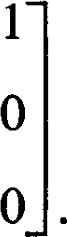
PROOF. Consider = 02T, which maps U into V. The kernel Ofc is T-1 (Vo). By assumption, I-Jo c T—I (VO). Hence, by the mapping factor theorem, there is a linear transformation i such that =

We say that is induced by T.

Numerical calculations with quotient spaces can usually be avoided in problems involving finite dimensional vector spaces. If U is a vector space over F and K is a subspace of U, we know from Theorem 4.9 of Chapter I that K is a direct summand. Let U = K O W. Then the canonical mapping  maps W isomorphically onto U/K. Thus any calculation involving U/K can be carried out in W.

Although there are many possible choices for the complementary subspace W, the Hermite normal form provides a simple and effective way to select a W and a basis for it. This typically arises in connection with a linear problem. To see this, reexamine the proof of Theorem 5.1. There we let k k be those indices for which c(4i)  We showed there that {F' K} where ß'i = formed a basis of c(U). {u oc } is a basis for a suitable W which is complementary to KG).

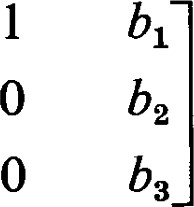
Example. Consider the linear transformation c of R5 into R3 represented by the matrix

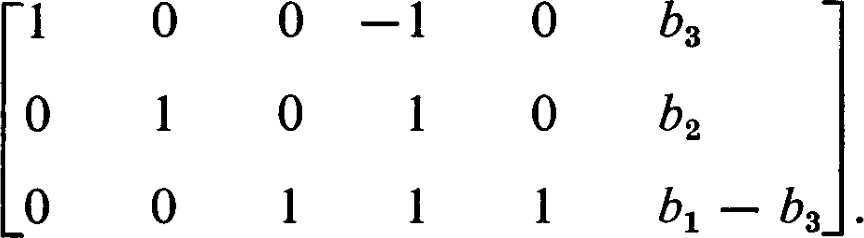
 1 O 1 o

O 1 o 1

1 

It is easy to determine that the kernel K of is 2-dimensional with basis 1, —l, l , O), (O, O, —l, O, l)}. This means that has rank 3 and the image of is all of R3. Thus R5 = R5/K is isomorphic to R3.

Consider the problem of solving the equation c(é) = F, where ß is represented by (bl, b2, b3). To solve this problem we reduce the augmented matrix 1 0 1 0 o 1 0 1 1 0 to the Hermite normal form

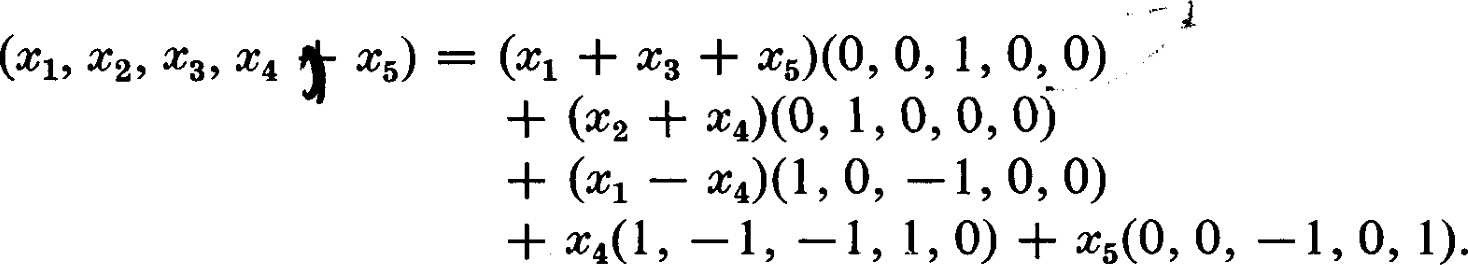


3

This means the solution is represented by (b3, b2, bl —

(h, b2, bl — b3, O, 0) = bl(0, O, 1, O, 0) + b2(0, 1, O, O, 0) + b3(1, O, —1, O, 0) is a particular solution and a convenient basis for a subspace W complementary to K is O, 1, O, 0), (0, 1, O, O, 0), (1 , 0, —l, 0, O)}. maps bl(0, O, 1, O, 0) + b2(0, l, O, O, 0) + b3(1 , 0, —1, 0, 0) onto (bl, b2, h). Hence, W is mapped isomorphically onto R3.

This example also provides an opportunity to illustrate the working of the first homomorphism theorem. For any (Xl, x2, $3, x4, $5) e R5.



Thus ($1, x2, 4, x4, x5) is mapped onto the coset (Xl + + O, l, (), 0) + — 0, —l, O, O) + K under the natural homomorphism onto R5/K. This coset is then mapped isomorphically onto — my) e R3. However, it is somewhat contrived to 11 | Hom(U, V) 83

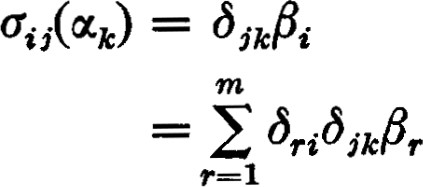
work out an example of this type. The main importance of the first homomorphism theorem is theoretical and not computational.

##### \*11 1 Hom(U, V)

Let U and V be vector spaces over F. We have already observed in Section I that the set of all linear transformations of U into V can be made into a vector space over F by defining addition and scalar multiplication appropriately. In this section we will explore some of the elementary consequences of this observation. We shall call this vector space Hom(U, V), "The space of all homomorphisms of U into V."

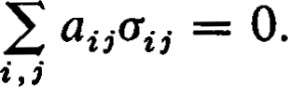
Theorem 11.1. If dim U = n and dim V = m, then dim Hom(U, V) = mn.

PROOF. Let {OCI, . . . , ocn} be a basis of U and let {ß} be a basis of V. Define the linear transformation of by the rule

 (11.1)

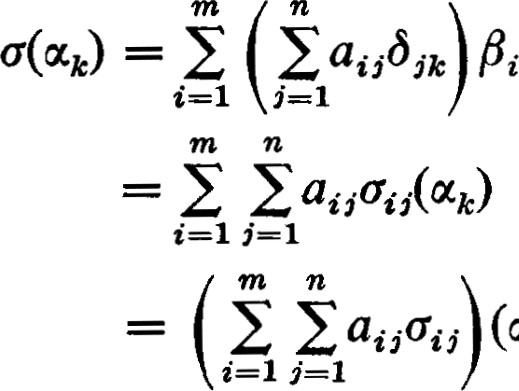
Thus is represented by the matrix [öriöik] = Aii. Aii has a zero in every position except for a I in row i column j.

The set {ct.,} is linearly independent. For if a linear relation existed among the it would be of the form



This means aijüii(æk) = 0 for all But auffii(otk) = aijöikÅi aikßt = O. Since {Pi} is a lineary independent set, aik —— 0 for 1 2 m. Since this is true for each k, all ail = O and {ct.,} is linearly independent.

If e Hom(U, V) and c(otk) = a ik 1Bi then

(otk).

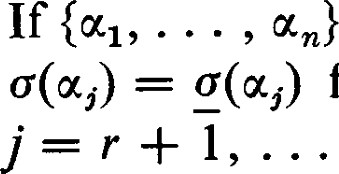
Thus {ct.j} spans Hom(U, V), which is therefore of dimension mn. 

If VI is a subspace of V, every linear transformation of U into VI also defines a mapping of U into V. This mapping of U into Vis a linear transformation of U into V. Thus, with each element of Hom(U, VI) there is associated in a natural way an element of Hom(U, V). We can identify Hom(U, VI) with a subset of Hom(U, V). With this identification Hom(U, VI) is a subspace of Hom(U, V).

Now let UI be a subspace of U. In this case we cannot consider Hom(U1, V) to be a subset of Hom(U, V) since a linear transformation in Hom(U1, V) is not necessarily defined on all of U. But any linear transformation in Hom(U, V) is certainly defined on UI. If c e Hom(U, V) we shall consider the mapping obtained by applying c only to elements in UI to be a new function and denote it by R(c). R(Ü) is called the restriction of c to UI. We can consider R(c) to be an element of Hom(U1, V).

It may happen that different linear transformations defined on U produce the same restriction on UI. We say that and are equivalent on UI if and only if R(ÜI) = R(Ü2). It is clear that + T) = R(c) + R(T) and R(ac) = aR(c) so that the mapping of Hom(U, V) into Hom(U1, V) is linear. We call this mapping R, the restriction mapping.

The kernel of R is clearly the set of all linear transformations in Hom(U, V) that vanish on UI. Let us denote this kernel by UI\*.

If is any linear transformation belonging to Hom(U1, V), it can be extended to a linear transformation belonging to Hom(U, V) in many ways. is a basis of U such that {al, oc } is a basis of UI, then let for J r, and let c(æj) be defined arbitrarily for n. Since is then the restriction of c, we see that R is an epimorphism of Hom(U, V) onto Hom(U1, V). Since Hom(U, V) is of dimension mn and Hom(U1, V) is of dimension mr, UI\* is of dimension m(n r). Theorem 11.2. Hom(U1, V) is canonically isomorphic to Hom(U, 

Note: It helps the intuitive understanding of this theorem to examine the method by which we obtained an extension of on UI, to on U. U\* is the set of all extensions of when is the zero mapping, and one can see directly that the dimension of is (n — r)m.