Chapter 5

Linear Algebra

The exalted position held by linear algebra is based upon the subject's ubiquitous utility and ease of application. The basic theory is developed here in full generality, i.e., modules are defined over an arbitrary ring R and not just over a field. The elementary facts about cosets, quotients, and homomorphisms follow the same pattern as in the chapters on groups and rings. We give a simple proof that if R is a commutative ring and $f : R^n \to R^n$ is a surjective R-module homomorphism, then f is an isomorphism. This shows that finitely generated free R-modules have a well defined dimension, and simplifies some of the development of linear algebra. It is in this chapter that the concepts about functions, solutions of equations, matrices, and generating sets come together in one unified theory.

After the general theory, we restrict our attention to vector spaces, i.e., modules over a field. The key theorem is that any vector space V has a free basis, and thus if V is finitely generated, it has a well defined dimension, and incredible as it may seem, this single integer determines V up to isomorphism. Also any endomorphism $f : V \to V$ may be represented by a matrix, and any change of basis corresponds to conjugation of that matrix. One of the goals in linear algebra is to select a basis so that the matrix representing f has a simple form. For example, if f is not injective, then f may be represented by a matrix whose first column is zero. As another example, if f is nilpotent, then f may be represented by a strictly upper triangular matrix. The theorem on Jordan canonical form is not proved in this chapter, and should not be considered part of this chapter. It is stated here in full generality only for reference and completeness. The proof is given in the Appendix. This chapter concludes with the study of real inner product spaces, and with the beautiful theory relating orthogonal matrices and symmetric matrices.

Definition Suppose R is a ring and M is an additive abelian group. The statement that M is a *right R-module* means there is a scalar multiplication

$$\begin{array}{lll} M \times R \to M & \text{satisfying} & (a_1 + a_2)r = a_1 r + a_2 r \\ (m, r) \to mr & & a(r_1 + r_2) = ar_1 + ar_2 \\ & & a(r_1 \cdot r_2) = (ar_1)r_2 \\ & & a\underline{1} = a \end{array}$$

for all $a, a_1, a_2 \in M$ and $r, r_1, r_2 \in R$.

The statement that M is a *left R-module* means there is a scalar multiplication

$$\begin{array}{lll} R \times M \to M & \text{satisfying} & r(a_1 + a_2) = ra_1 + ra_2 \\ (r, m) \to rm & & (r_1 + r_2)a = r_1 a + r_2 a \\ & & (r_1 \cdot r_2)a = r_1(r_2 a) \\ & & \underline{1}a = a \end{array}$$

Note that the plus sign is used ambiguously, as addition in M and as addition in R.

Notation The fact that M is a right (left) R-module will be denoted by $M = M_R$ ($M = {}_R M$). If R is commutative and $M = M_R$ then left scalar multiplication defined by $ra = ar$ makes M into a left R-module. Thus for commutative rings, we may write the scalars on either side. In this text we stick to right R-modules.

Convention Unless otherwise stated, it is assumed that R is a ring and the word "R-module" (or sometimes just "module") means "right R-module".

Theorem Suppose M is an R-module.

1) If $r \in R$, then $f : M \to M$ defined by $f(a) = ar$ is a homomorphism of additive groups. In particular $(\underline{0}_M)r = \underline{0}_M$.
2) If $a \in M$, $a\underline{0}_R = \underline{0}_M$.
3) If $a \in M$ and $r \in R$, then $(-a)r = -(ar) = a(-r)$.

Proof This is a good exercise in using the axioms for an R-module.

Submodules If M is an R-module, the statement that a subset $N \subset M$ is a *submodule* means it is a subgroup which is closed under scalar multiplication, i.e., if $a \in N$ and $r \in R$, then $ar \in N$. In this case N will be an R-module because the axioms will automatically be satisfied. Note that $\underline{0}$ and M are submodules, called the *improper* submodules of M.

Theorem Suppose M is an R-module, T is an index set, and for each $t \in T$, N_t is a submodule of M.

1) $\bigcap_{t\in T} N_t$ is a submodule of M.
2) If $\{N_t\}$ is a monotonic collection, $\bigcup_{t\in T} N_t$ is a submodule.
3) $+_{t\in T} N_t = \{$all finite sums $a_1 + \cdots + a_m$: each a_i belongs to some $N_t\}$ is a submodule. If $T = \{1, 2, .., n\}$, then this submodule may be written as $N_1 + N_2 + \cdots + N_n = \{a_1 + a_2 + \cdots + a_n : \text{ each } a_i \in N_i\}$.

Proof We know from page 22 that versions of 1) and 2) hold for subgroups, and in particular for subgroups of additive abelian groups. To finish the proofs it is only necessary to check scalar multiplication, which is immediate. Also the proof of 3) is immediate. Note that if N_1 and N_2 are submodules of M, $N_1 + N_2$ is the smallest submodule of M containing $N_1 \cup N_2$.

Exercise Suppose T is a non-void set, N is an R-module, and N^T is the collection of all functions $f : T \to N$ with addition defined by $(f+g)(t) = f(t)+g(t)$, and scalar multiplication defined by $(fr)(t) = f(t)r$. Show N^T is an R-module. (We know from the last exercise in Chapter 2 that N^T is a group, and so it is only necessary to check scalar multiplication.) This simple fact is quite useful in linear algebra. For example, in 5) of the theorem below, it is stated that $\mathrm{Hom}_R(M, N)$ forms an abelian group. So it is only necessary to show that $\mathrm{Hom}_R(M, N)$ is a subgroup of N^M. Also in 8) it is only necessary to show that $\mathrm{Hom}_R(M, N)$ is a submodule of N^M.

Homomorphisms

Suppose M and N are R-modules. A function $f : M \to N$ is a *homomorphism* (i.e., an R-module homomorphism) provided it is a group homomorphism and if $a \in M$ and $r \in R$, $f(ar) = f(a)r$. On the left, scalar multiplication is in M and on the right it is in N. The basic facts about homomorphisms are listed below. Much

of this work has already been done in the chapter on groups (see page 28).

Theorem

1) The zero map $M \to N$ is a homomorphism.

2) The identity map $I : M \to M$ is a homomorphism.

3) The composition of homomorphisms is a homomorphism.

4) The sum of homomorphisms is a homomorphism. If $f, g : M \to N$ are homomorphisms, define $(f+g) : M \to N$ by $(f+g)(a) = f(a) + g(a)$. Then $f+g$ is a homomorphism. Also $(-f)$ defined by $(-f)(a) = -f(a)$ is a homomorphism. If $h : N \to P$ is a homomorphism, $h \circ (f+g) = (h \circ f) + (h \circ g)$. If $k : P \to M$ is a homomorphism, $(f+g) \circ k = (f \circ k) + (g \circ k)$.

5) $\mathrm{Hom}_R(M, N) = \mathrm{Hom}(M_R, N_R)$, the set of all homomorphisms from M to N, forms an abelian group under addition. $\mathrm{Hom}_R(M, M)$, with multiplication defined to be composition, is a ring.

6) If a bijection $f : M \to N$ is a homomorphism, then $f^{-1} : N \to M$ is also a homomorphism. In this case f and f^{-1} are called *isomorphisms*. A homomorphism $f : M \to M$ is called an *endomorphism*. An isomorphism $f : M \to M$ is called an *automorphism*. The units of the endomorphism ring $\mathrm{Hom}_R(M, M)$ are the automorphisms. Thus the automorphisms on M form a group under composition. We will see later that if $M = R^n$, $\mathrm{Hom}_R(R^n, R^n)$ is just the matrix ring R_n and the automorphisms are merely the invertible matrices.

7) If R is commutative and $r \in R$, then $g : M \to M$ defined by $g(a) = ar$ is a homomorphism. Furthermore, if $f : M \to N$ is a homomorphism, fr defined by $(fr)(a) = f(ar) = f(a)r$ is a homomorphism.

8) If R is commutative, $\mathrm{Hom}_R(M, N)$ is an R-module.

9) Suppose $f : M \to N$ is a homomorphism, $G \subset M$ is a submodule, and $H \subset N$ is a submodule. Then $f(G)$ is a submodule of N and $f^{-1}(H)$ is a submodule of M. In particular, image(f) is a submodule of N and $\ker(f) = f^{-1}(\underline{0})$ is a submodule of M.

Proof This is just a series of observations.

Abelian groups are Z-modules On page 21, it is shown that any additive group M admits a scalar multiplication by integers, and if M is abelian, the properties are satisfied to make M a **Z**-module. Note that this is the only way M can be a **Z**-module, because $a1 = a$, $a2 = a + a$, etc. Furthermore, if $f : M \to N$ is a group homomorphism of abelian groups, then f is also a **Z**-module homomorphism.

Summary Additive abelian groups are "the same things" as **Z**-modules. While group theory in general is quite separate from linear algebra, the study of additive abelian groups is a special case of the study of R-modules.

Exercise R-modules are also **Z**-modules and R-module homomorphisms are also **Z**-module homomorphisms. If M and N are **Q**-modules and $f : M \to N$ is a **Z**-module homomorphism, must it also be a **Q**-module homomorphism?

Homomorphisms on R^n

R^n as an R-module On page 54 it was shown that the additive abelian group $R_{m,n}$ admits a scalar multiplication by elements in R. The properties listed there were exactly those needed to make $R_{m,n}$ an R-module. Of particular importance is the case $R^n = R \oplus \cdots \oplus R = R_{n,1}$ (see page 53). We begin with the case $n = 1$.

R as a right R-module Let $M = R$ and define scalar multiplication on the right by $ar = a \cdot r$. That is, scalar multiplication is just ring multiplication. This makes R a right R-module denoted by R_R (or just R). This is the same as the definition before for R^n when $n = 1$.

Theorem Suppose R is a ring and N is a subset of R. Then N is a submodule of R_R ($_RR$) iff N is a right (left) ideal of R.

Proof The definitions are the same except expressed in different language.

Theorem Suppose $M = M_R$ and $f, g : R \to M$ are homomorphisms with $f(\underline{1}) = g(\underline{1})$. Then $f = g$. Furthermore, if $m \in M$, $\exists!$ homomorphism $h : R \to M$ with $h(\underline{1}) = m$. In other words, $\mathrm{Hom}_R(R, M) \approx M$.

Proof Suppose $f(\underline{1}) = g(\underline{1})$. Then $f(r) = f(\underline{1} \cdot r) = f(\underline{1})r = g(\underline{1})r = g(\underline{1} \cdot r) = g(r)$. Given $m \in M$, $h : R \to M$ defined by $h(r) = mr$ is a homomorphism. Thus

evaluation at $\underline{1}$ gives a bijection from $\mathrm{Hom}_R(R, M)$ to M, and this bijection is clearly a group isomorphism. If R is commutative, it is an isomorphism of R-modules.

In the case $M = R$, the above theorem states that multiplication on left by some $m \in R$ defines a right R-module homomorphism from R to R, and every module homomorphism is of this form. The element m should be thought of as a 1×1 matrix. We now consider the case where the domain is R^n.

Homomorphisms on R^n Define $e_i \in R^n$ by $e_i = \begin{pmatrix} \underline{0} \\ \cdot \\ \underline{1}_i \\ \cdot \\ \underline{0} \end{pmatrix}$. Note that any $\begin{pmatrix} r_1 \\ \cdot \\ \\ \cdot \\ r_n \end{pmatrix}$ can be written uniquely as $e_1 r_1 + \cdots + e_n r_n$. The sequence $\{e_1, .., e_n\}$ is called the *canonical free basis* or *standard basis* for R^n.

Theorem Suppose $M = M_R$ and $f, g : R^n \to M$ are homomorphisms with $f(e_i) = g(e_i)$ for $1 \leq i \leq n$. Then $f = g$. Furthermore, if $m_1, m_2, ..., m_n \in M$, $\exists!$ homomorphism $h : R^n \to M$ with $h(e_i) = m_i$ for $1 \leq i \leq m$. The homomorphism h is defined by $h(e_1 r_1 + \cdots + e_n r_n) = m_1 r_1 + \cdots + m_n r_n$.

Proof The proof is straightforward. Note this theorem gives a bijection from $\mathrm{Hom}_R(R^n, M)$ to $M^n = M \times M \times \cdots \times M$ and this bijection is a group isomorphism. We will see later that the product M^n is an R-module with scalar multiplication defined by $(m_1, m_2, .., m_n)r = (m_1 r, m_2 r, .., m_n r)$. If R is commutative so that $\mathrm{Hom}_R(R^n, M)$ is an R-module, this theorem gives an R-module isomorphism from $\mathrm{Hom}_R(R^n, M)$ to M^n.

This theorem reveals some of the great simplicity of linear algebra. It does not matter how complicated the ring R is, or which R-module M is selected. Any R-module homomorphism from R^n to M is determined by its values on the basis, and any function from that basis to M extends uniquely to a homomorphism from R^n to M.

Exercise Suppose R is a field and $f : R_R \to M$ is a non-zero homomorphism. Show f is injective.

Now let's examine the special case $M = R^m$ and show $\mathrm{Hom}_R(R^n, R^m) \approx R_{m,n}$.

Theorem Suppose $A = (a_{i,j}) \in R_{m,n}$. Then $f : R^n \to R^m$ defined by $f(B) = AB$ is a homomorphism with $f(e_i) =$ column i of A. Conversely, if $v_1, \ldots, v_n \in R^m$, define $A \in R_{m,n}$ to be the matrix with column $i = v_i$. Then f defined by $f(B) = AB$ is the unique homomorphism from R^n to R^m with $f(e_i) = v_i$.

Even though this follows easily from the previous theorem and properties of matrices, it is one of the great classical facts of linear algebra. Matrices over R give R-module homomorphisms! Furthermore, addition of matrices corresponds to addition of homomorphisms, and multiplication of matrices corresponds to composition of homomorphisms. These properties are made explicit in the next two theorems.

Theorem If $f, g : R^n \to R^m$ are given by matrices $A, C \in R_{m,n}$, then $f + g$ is given by the matrix $A+C$. Thus $\mathrm{Hom}_R(R^n, R^m)$ and $R_{m,n}$ are isomorphic as additive groups. If R is commutative, they are isomorphic as R-modules.

Theorem If $f\colon R^n \to R^m$ is the homomorphism given by $A \in R_{m,n}$ and $g : R^m \to R^p$ is the homomorphism given by $C \in R_{p,m}$, then $g \circ f : R^n \to R^p$ is given by $CA \in R_{p,n}$. That is, composition of homomorphisms corresponds to multiplication of matrices.

Proof This is just the associative law of matrix multiplication, $C(AB) = (CA)B$.

The previous theorem reveals where matrix multiplication comes from. It is the matrix which represents the composition of the functions. In the case where the domain and range are the same, we have the following elegant corollary.

Corollary $\mathrm{Hom}_R(R^n, R^n)$ and R_n are isomorphic as rings. The automorphisms correspond to the invertible matrices.

This corollary shows one way non-commutative rings arise, namely as endomorphism rings. Even if R is commutative, R_n is never commutative unless $n = 1$.

We now return to the general theory of modules (over some given ring R).

Cosets and Quotient Modules

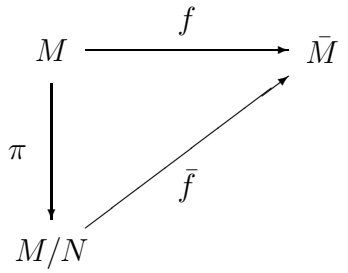
After seeing quotient groups and quotient rings, quotient modules go through without a hitch. As before, R is a ring and module means R-module.

Theorem Suppose M is a module and $N \subset M$ is a submodule. Since N is a normal subgroup of M, the additive abelian quotient group M/N is defined. Scalar multiplication defined by $(a + N)r = (ar + N)$ is well defined and gives M/N the structure of an R-module. The natural projection $\pi : M \to M/N$ is a surjective homomorphism with kernel N. Furthermore, if $f : M \to \bar{M}$ is a surjective homomorphism with $\ker(f) = N$, then $M/N \approx \bar{M}$ (see below).

Proof On the group level, this is all known from Chapter 2 (see pages 27 and 29). It is only necessary to check the scalar multiplication, which is obvious.

The relationship between quotients and homomorphisms for modules is the same as for groups and rings, as shown by the next theorem.

Theorem Suppose $f : M \to \bar{M}$ is a homomorphism and N is a submodule of M. If $N \subset \ker(f)$, then $\bar{f} : (M/N) \to \bar{M}$ defined by $\bar{f}(a + N) = f(a)$ is a well defined homomorphism making the following diagram commute.



Thus defining a homomorphism on a quotient module is the same as defining a homomorphism on the numerator that sends the denominator to $\underline{0}$. The image of $\bar{f}$ is the image of f, and the kernel of $\bar{f}$ is $\ker(f)/N$. Thus if $N = \ker(f)$, $\bar{f}$ is injective, and thus $(M/N) \approx \text{image}(f)$. Therefore for any homomorphism f, $(\text{domain}(f)/\ker(f)) \approx \text{image}(f)$.

Proof On the group level this is all known from Chapter 2 (see page 29). It is only necessary to check that $\bar{f}$ is a module homomorphism, and this is immediate.

Theorem Suppose M is an R-module and K and L are submodules of M.

i) The natural homomorphism $K \to (K+L)/L$ is surjective with kernel $K \cap L$. Thus $(K/K \cap L) \xrightarrow{\approx} (K+L)/L$ is an isomorphism.

ii) Suppose $K \subset L$. The natural homomorphism $M/K \to M/L$ is surjective with kernel L/K. Thus $(M/K)/(L/K) \xrightarrow{\approx} M/L$ is an isomorphism.

Examples These two examples are for the case $R = \mathbf{Z}$, i.e., for abelian groups.

1) $M = \mathbf{Z} \quad K = 3\mathbf{Z} \quad L = 5\mathbf{Z} \quad K \cap L = 15\mathbf{Z} \quad K + L = \mathbf{Z}$
$K/K \cap L = 3\mathbf{Z}/15\mathbf{Z} \approx \mathbf{Z}/5\mathbf{Z} = (K+L)/L$

2) $M = \mathbf{Z} \quad K = 6\mathbf{Z} \quad L = 3\mathbf{Z} \quad (K \subset L)$
$(M/K)/(L/K) = (\mathbf{Z}/6\mathbf{Z})/(3\mathbf{Z}/6\mathbf{Z}) \approx \mathbf{Z}/3\mathbf{Z} = M/L$

Products and Coproducts

Infinite products work fine for modules, just as they do for groups and rings. This is stated below in full generality, although the student should think of the finite case. In the finite case something important holds for modules that does not hold for non-abelian groups or rings – namely, the finite product is also a coproduct. This makes the structure of module homomorphisms much more simple. For the finite case we may use either the product or sum notation, i.e., $M_1 \times M_2 \times \cdots \times M_n = M_1 \oplus M_2 \oplus \cdots \oplus M_n$.

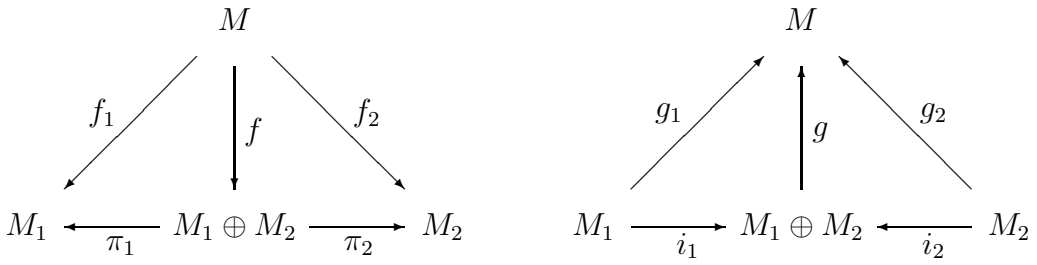
Theorem Suppose T is an index set and for each $t \in T$, M_t is an R-module. On the additive abelian group $\prod_{t \in T} M_t = \prod M_t$ define scalar multiplication by $\{m_t\}r = \{m_t r\}$. Then $\prod M_t$ is an R-module and, for each $s \in T$, the natural projection $\pi_s : \prod M_t \to M_s$ is a homomorphism. Suppose M is a module. Under the natural 1-1 correspondence from $\{\text{functions } f : M \to \prod M_t\}$ to $\{\text{sequence of functions } \{f_t\}_{t \in T}$ where $f_t : M \to M_t\}$, f is a homomorphism iff each f_t is a homomorphism.

Proof We already know from Chapter 2 that f is a group homomorphism iff each f_t is a group homomorphism. Since scalar multiplication is defined coordinatewise, f is a module homomorphism iff each f_t is a module homomorphism.

Definition If T is finite, the coproduct and product are the same module. If T is infinite, the *coproduct* or *sum* $\coprod_{t\in T} M_t = \bigoplus_{t\in T} M_t = \oplus M_t$ is the submodule of $\prod M_t$ consisting of all sequences $\{m_t\}$ with only a finite number of non-zero terms. For each $s \in T$, the inclusion homomorphisms $i_s : M_s \to \oplus M_t$ is defined by $i_s(a) = \{a_t\}$ where $a_t = \underline{0}$ if $t \neq s$ and $a_s = a$. Thus each M_s may be considered to be a submodule of $\oplus M_t$.

Theorem Suppose M is an R-module. There is a 1-1 correspondence from $\{$homomorphisms $g : \oplus M_t \to M\}$ and $\{$sequences of homomorphisms $\{g_t\}_{t\in T}$ where $g_t : M_t \to M\}$. Given g, g_t is defined by $g_t = g \circ i_t$. Given $\{g_t\}$, g is defined by $g(\{m_t\}) = \sum_t g_t(m_t)$. Since there are only a finite number of non-zero terms, this sum is well defined.

For $T = \{1, 2\}$ the product and sum properties are displayed in the following commutative diagrams.



Theorem For finite T, the 1-1 correspondences in the above theorems actually produce group isomorphisms. If R is commutative, they give isomorphisms of R-modules.

$$\mathrm{Hom}_R(M, M_1 \oplus \cdots \oplus M_n) \approx \mathrm{Hom}_R(M, M_1) \oplus \cdots \oplus \mathrm{Hom}_R(M, M_n) \quad \text{and}$$
$$\mathrm{Hom}_R(M_1 \oplus \cdots \oplus M_n, M) \approx \mathrm{Hom}_R(M_1, M) \oplus \cdots \oplus \mathrm{Hom}_R(M_n, M)$$

Proof Let's look at this theorem for products with $n = 2$. All it says is that if $f = (f_1, f_2)$ and $h = (h_1, h_2)$, then $f + h = (f_1 + h_1, f_2 + h_2)$. If R is commutative, so that the objects are R-modules and not merely additive groups, then the isomorphisms are module isomorphisms. This says merely that $fr = (f_1, f_2)r = (f_1 r, f_2 r)$.

Exercise Suppose M and N are R-modules. Show that $M \oplus N$ is isomorphic to $N \oplus M$. Now suppose $A \subset M,\ B \subset N$ are submodules and show $(M \oplus N)/(A \oplus B)$ is isomorphic to $(M/A) \oplus (N/B)$. In particular, if $a \in R$ and $b \in R$, then $(R \oplus R)/(aR \oplus bR)$ is isomorphic to $(R/aR) \oplus (R/bR)$. For example, the abelian group $(\mathbf{Z} \oplus \mathbf{Z})/(2\mathbf{Z} \oplus 3\mathbf{Z})$ is isomorphic to $\mathbf{Z}_2 \oplus \mathbf{Z}_3$. These isomorphisms are transparent and are used routinely in algebra without comment (see Th 4, page 118).

Exercise Suppose R is a commutative ring, M is an R-module, and $n \geq 1$. Define a function $\alpha : \mathrm{Hom}_R(R^n, M) \to M^n$ which is a R-module isomorphism.

Summands

One basic question in algebra is "When does a module split as the sum of two modules?". Before defining summand, here are two theorems for background.

Theorem Consider $M_1 = M_1 \oplus \underline{0}$ as a submodule of $M_1 \oplus M_2$. Then the projection map $\pi_2 : M_1 \oplus M_2 \to M_2$ is a surjective homomorphism with kernel M_1. Thus $(M_1 \oplus M_2)/M_1$ is isomorphic to M_2. (See page 35 for the group version.)

This is exactly what you would expect, and the next theorem is almost as intuitive.

Theorem Suppose K and L are submodules of M and $f : K \oplus L \to M$ is the natural homomorphism, $f(k, l) = k + l$. Then the image of f is $K + L$ and the kernel of f is $\{(a, -a) : a \in K \cap L\}$. Thus f is an isomorphism iff $K + L = M$ and $K \cap L = \underline{0}$. In this case we write $K \oplus L = M$. This abuse of notation allows us to avoid talking about "internal" and "external" direct sums.

Definition Suppose K is a submodule of M. The statement that K is a *summand* of M means $\exists$ a submodule L of M with $K \oplus L = M$. According to the previous theorem, this is the same as there exists a submodule L with $K + L = M$ and $K \cap L = \underline{0}$. If such an L exists, it need not be unique, but it will be unique up to isomorphism, because $L \approx M/K$. Of course, M and $\underline{0}$ are always summands of M.

Exercise Suppose M is a module and $K = \{(m, m) : m \in M\} \subset M \oplus M$. Show K is a submodule of $M \oplus M$ which is a summand.

Exercise $\mathbf{R}$ is a module over $\mathbf{Q}$, and $\mathbf{Q} \subset \mathbf{R}$ is a submodule. Is $\mathbf{Q}$ a summand of $\mathbf{R}$? With the material at hand, this is not an easy question. Later on, it will be easy.

Exercise Answer the following questions about abelian groups, i.e., **Z**-modules. (See the third exercise on page 35.)

1) Is $2\mathbf{Z}$ a summand of $\mathbf{Z}$?
2) Is $2\mathbf{Z}_4$ a summand of $\mathbf{Z}_4$?
3) Is $3\mathbf{Z}_{12}$ a summand of $\mathbf{Z}_{12}$?
4) Suppose $m, n > 1$. When is $n\mathbf{Z}_{mn}$ a summand of $\mathbf{Z}_{mn}$?

Exercise If T is a ring, define the center of T to be the subring $\{t : ts = st \text{ for all } s \in T\}$. Let R be a commutative ring and $T = R_n$. There is a exercise on page 57 to show that the center of T is the subring of scalar matrices. Show R^n is a left T-module and find $\mathrm{Hom}_T(R^n, R^n)$.

Independence, Generating Sets, and Free Basis

This section is a generalization and abstraction of the brief section **Homomorphisms on** R^n. These concepts work fine for an infinite index set T because linear combination means finite linear combination. However, to avoid dizziness, the student should first consider the case where T is finite.

Definition Suppose M is an R-module, T is an index set, and for each $t \in T$, $s_t \in M$. Let S be the sequence $\{s_t\}_{t\in T} = \{s_t\}$. The statement that S is *dependent* means $\exists$ a finite number of distinct elements $t_1, ..., t_n$ in T, and elements $r_1, .., r_n$ in R, not all zero, such that the linear combination $s_{t_1}r_1 + \cdots + s_{t_n}r_n = \underline{0}$. Otherwise, S is *independent*. Note that if some $s_t = \underline{0}$, then S is dependent. Also if $\exists$ distinct elements t_1 and t_2 in T with $s_{t_1} = s_{t_2}$, then S is dependent.

Let SR be the set of all linear combinations $s_{t_1}r_1 + \cdots + s_{t_n}r_n$. SR is a submodule of M called the submodule *generated* by S. If S is independent and generates M, then S is said to be a *basis* or *free basis* for M. In this case any $v \in M$ can be written uniquely as a linear combination of elements in S. An R-module M is said to be a free R-module if it is zero or if it has a basis. The next two theorems are obvious, except for the confusing notation. You might try first the case $T = \{1, 2, ..., n\}$ and $\oplus R_t = R^n$ (see p 72).

Theorem For each $t \in T$, let $R_t = R_R$ and for each $c \in T$, let $e_c \in \oplus R_t = \bigoplus_{t\in T} R_t$ be $e_c = \{r_t\}$ where $r_c = \underline{1}$ and $r_t = \underline{0}$ if $t \neq c$. Then $\{e_c\}_{c\in T}$ is a basis for $\oplus R_t$ called the *canonical basis* or *standard basis*.

Theorem Suppose N is an R-module and M is a free R-module with a basis $\{s_t\}$. Then $\exists$ a 1-1 correspondence from the set of all functions $g:\{s_t\} \to N$ and the set of all homomorphisms $f: M \to N$. Given g, define f by $f(s_{t_1}r_1 + \cdots + s_{t_n}r_n) = g(s_{t_1})r_1 + \cdots + g(s_{t_n})r_n$. Given f, define g by $g(s_t) = f(s_t)$. In other words, f is completely determined by what it does on the basis S, and you are "free" to send the basis any place you wish and extend to a homomorphism.

Recall that we have already had the preceding theorem in the case S is the canonical basis for $M = R^n$ (p 72). The next theorem is so basic in linear algebra that it is used without comment. Although the proof is easy, it should be worked carefully.

Theorem Suppose N is a module, M is a free module with basis $S = \{s_t\}$, and $f: M \to N$ is a homomorphism. Let $f(S)$ be the sequence $\{f(s_t)\}$ in N.

1) $f(S)$ generates N iff f is surjective.
2) $f(S)$ is independent in N iff f is injective.
3) $f(S)$ is a basis for N iff f is an isomorphism.
4) If $h: M \to N$ is a homomorphism, then $f = h$ iff $f \mid S = h \mid S$.

Exercise Let $(A_1, .., A_n)$ be a sequence of n vectors with each $A_i \in \mathbf{Z}^n$. Show this sequence is linearly independent over $\mathbf{Z}$ iff it is linearly independent over $\mathbf{Q}$. Is it true the sequence is linearly independent over $\mathbf{Z}$ iff it is linearly independent over $\mathbf{R}$? This question is difficult until we learn more linear algebra.

Characterization of Free Modules

Any free R-module is isomorphic to one of the canonical free R-modules $\oplus R_t$. This is just an observation, but it is a central fact in linear algebra.

Theorem A non-zero R-module M is free iff $\exists$ an index set T such that $M \approx \bigoplus_{t \in T} R_t$. In particular, M has a finite free basis of n elements iff $M \approx R^n$.

Proof If M is isomorphic to $\oplus R_t$ then M is certainly free. So now suppose M has a free basis $\{s_t\}$. Then the homomorphism $f: M \to \oplus R_t$ with $f(s_t) = e_t$ sends the basis for M to the canonical basis for $\oplus R_t$. By 3) in the preceding theorem, f is an isomorphism.

Exercise Suppose R is a commutative ring, $A \in R_n$, and the homomorphism $f : R^n \to R^n$ defined by $f(B) = AB$ is surjective. Show f is an isomorphism, i.e., show A is invertible. This is a key theorem in linear algebra, although it is usually stated only for the case where R is a field. Use the fact that $\{e_1, .., e_n\}$ is a free basis for R^n.

The next exercise is routine, but still informative.

Exercise Let $R = \mathbf{Z}$, $A = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 2 & -5 \end{pmatrix}$ and f: $\mathbf{Z}^3 \to \mathbf{Z}^2$ be the group homomorphism defined by A. Find a non-trivial linear combination of the columns of A which is $\underline{0}$. Also find a non-zero element of kernel(f).

If R is a commutative ring, you can relate properties of R as an R-module to properties of R as a ring.

Exercise Suppose R is a commutative ring and $v \in R$, $v \neq \underline{0}$.

1) v is independent iff v is _____.
2) v is a basis for R iff v generates R iff v is _____.

Note that 2) here is essentially the first exercise for the case $n = 1$. That is, if $f : R \to R$ is a surjective R-module homomorphism, then f is an isomorphism.

Relating these concepts to matrices

The theorem stated below gives a summary of results we have already had. It shows that certain concepts about matrices, linear independence, injective homomorphisms, and solutions of equations, are all the same — they are merely stated in different language. Suppose $A \in R_{m,n}$ and $f : R^n \to R^m$ is the homomorphism associated with A, i.e., $f(B) = AB$. Let $v_1, .., v_n \in R^m$ be the columns of A, i.e., $f(e_i) = v_i$ = column i of A. Let $B = \begin{pmatrix} b_1 \\ . \\ b_n \end{pmatrix}$ represent an element of R^n and $C = \begin{pmatrix} c_1 \\ . \\ c_m \end{pmatrix}$

represent an element of R^m.

Theorem

1) The element $f(B)$ is a linear combination of the columns of A, that is $f(B) = f(e_1b_1 + \cdots + e_nb_n) = v_1b_1 + \cdots + v_nb_n$. Thus the image of f is generated by the columns of A.

2) $\{v_1, .., v_n\}$ generates R^m iff f is surjective iff (for any $C \in R^m$, $AX = C$ has a solution).

3) $\{v_1, .., v_n\}$ is independent iff f is injective iff $AX = \underline{0}$ has a unique solution iff ($\exists$ $C \in R^m$ such that $AX = C$ has a unique solution).

4) $\{v_1, .., v_n\}$ is a basis for R^m iff f is an isomorphism iff (for any $C \in R^m$, $AX = C$ has a unique solution).

Relating these concepts to square matrices

We now look at the preceding theorem in the special case where $n = m$ and R is a commutative ring. So far in this chapter we have just been cataloging. Now we prove something more substantial, namely that if $f : R^n \to R^n$ is surjective, then f is injective. Later on we will prove that if R is a field, injective implies surjective.

Theorem Suppose R is a commutative ring, $A \in R_n$, and $f : R^n \to R^n$ is defined by $f(B) = AB$. Let $v_1, .., v_n \in R^n$ be the columns of A, and $w_1, .., w_n \in R^n = R_{1,n}$ be the rows of A. Then the following are equivalent.

1) f is an automorphism.

2) A is invertible, i.e., $\mid A \mid$ is a unit in R.

3) $\{v_1, .., v_n\}$ is a basis for R^n.

4) $\{v_1, .., v_n\}$ generates R^n.

5) f is surjective.

2^t) A^t is invertible, i.e., $\mid A^t \mid$ is a unit in R.

3^t) $\{w_1, .., w_n\}$ is a basis for R^n.

$4^t)$ $\{w_1, .., w_n\}$ generates R^n.

Proof Suppose 5) is true and show 2). Since f is onto, $\exists\ u_1, ..., u_n \in R^n$ with $f(u_i) = e_i$. Let $g : R^n \to R^n$ be the homomorphism satisfying $g(e_i) = u_i$. Then $f \circ g$ is the identity. Now g comes from some matrix D and thus $AD = I$. This shows that A has a right inverse and is thus invertible. Recall that the proof of this fact uses determinant, which requires that R be commutative (see the exercise on page 64).

We already know the first three properties are equivalent, 4) and 5) are equivalent, and 3) implies 4). Thus the first five are equivalent. Furthermore, applying this result to A^t shows that the last three properties are equivalent to each other. Since $\mid A \mid = \mid A^t \mid$, 2) and 2^t) are equivalent.

Uniqueness of Dimension

There exists a ring R with $R^2 \approx R^3$ as R-modules, but this is of little interest. If R is commutative, this is impossible, as shown below. First we make a convention.

Convention *For the remainder of this chapter, R will be a commutative ring.*

Theorem If $f : R^m \to R^n$ is a surjective R-module homomorphism, then $m \geq n$.

Proof Suppose $k = n - m$ is positive. Define $h : (R^m \oplus R^k = R^n) \to R^n$ by $h(u, v) = f(u)$. Then h is a surjective homomorphism, and by the previous section, also injective. This is a contradiction and thus $m \geq n$.

Corollary If $f : R^m \to R^n$ is an isomorphism, then $m = n$.

Proof Each of f and f^{-1} is surjective, so $m = n$ by the previous theorem.

Corollary If $\{v_1, .., v_m\}$ generates R^n, then $m \geq n$.

Proof The hypothesis implies there is a surjective homomorphism $R^m \to R^n$. So this follows from the first theorem.

Lemma Suppose M is a f.g. module (i.e., a finite generated R-module). Then if M has a basis, that basis is finite.

Proof Suppose $U \subset M$ is a finite generating set and S is a basis. Then any element of U is a finite linear combination of elements of S, and thus S is finite.

Theorem Suppose M is a f.g. module. If M has a basis, that basis is finite and any other basis has the same number of elements. This number is denoted by $\dim(M)$, the *dimension* of M. (By convention, $\underline{0}$ is a free module of dimension 0.)

Proof By the previous lemma, any basis for M must be finite. M has a basis of n elements iff $M \approx R^n$. The result follows because $R^n \approx R^m$ iff $n = m$.

Change of Basis

Before changing basis, we recall what a basis is. Previously we defined generating, independence, and basis for sequences, not for collections. For the concept of generating it matters not whether you use sequences or collections, but for independence and basis, you must use sequences. Consider the columns of the real matrix $A = \begin{pmatrix} 2 & 3 & 2 \\ 1 & 4 & 1 \end{pmatrix}$. If we consider the column vectors of A as a collection, there are only two of them, yet we certainly don't wish to say the columns of A form a basis for $\mathbf{R}^2$. In a set or collection, there is no concept of repetition. In order to make sense, we must consider the columns of A as an ordered triple of vectors, and this sequence is dependent. In the definition of basis on page 78, basis is defined for sequences, not for sets or collections.

Two sequences cannot begin to be equal unless they have the same index set. Here we follow the classical convention that an index set with n elements will be $\{1, 2, .., n\}$, and thus a basis for M with n elements is a sequence $S = \{u_1, .., u_n\}$ or if you wish, $S = (u_1, .., u_n) \in M^n$. Suppose M is an R-module with a basis of n elements. Recall there is a bijection $\alpha : \mathrm{Hom}_R(R^n, M) \to M^n$ defined by $\alpha(h) = (h(e_1), .., h(e_n))$. Now $h : R^n \to M$ is an isomorphism iff $\alpha(h)$ is a basis for M.

Summary The point of all this is that selecting a basis of n elements for M is the same as selecting an isomorphism from R^n to M, and from this viewpoint, change of basis can be displayed by the diagram below.

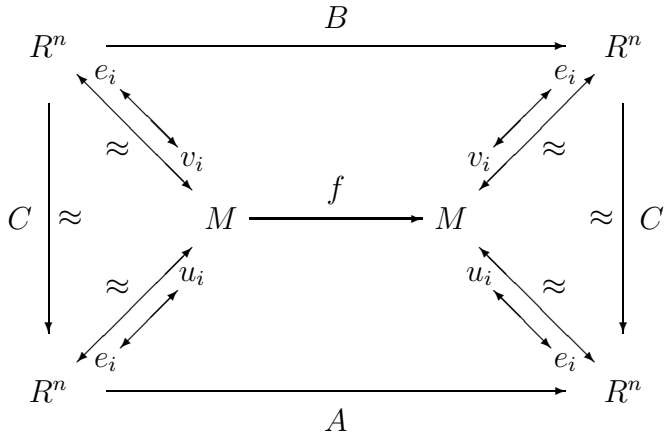
Endomorphisms on R^n are represented by square matrices, and thus have a determinant and trace. Now suppose M is a f.g. free module and $f : M \to M$ is a homomorphism. In order to represent f by a matrix, we must select a basis for M (i.e., an isomorphism with R^n). We will show that this matrix is well defined up to similarity, and thus the determinant, trace, and characteristic polynomial of f are well defined.

Definition Suppose M is a free module, $S = \{u_1, .., u_n\}$ is a basis for M, and $f : M \to M$ is a homomorphism. The matrix $A = (a_{i,j}) \in R_n$ of f w.r.t. the basis S is defined by $f(u_i) = u_1 a_{1,i} + \cdots + u_n a_{n,i}$. (Note that if $M = R^n$ and $u_i = e_i$, A is the usual matrix associated with f).

Theorem Suppose $T = \{v_1, .., v_n\}$ is another basis for M and $B \in R_n$ is the matrix of f w.r.t. T. Define $C = (c_{i,j}) \in R_n$ by $v_i = u_1 c_{1,i} + \cdots + u_n c_{n,i}$. Then C is invertible and $B = C^{-1}AC$, i.e., A and B are similar. Therefore $|A| = |B|$, trace(A)=trace(B), and A and B have the same characteristic polynomial (see page 66 of chapter 4).

Conversely, suppose $C = (c_{i,j}) \in R_n$ is invertible. Define $T = \{v_1, .., v_n\}$ by $v_i = u_1 c_{1,i} + \cdots + u_n c_{n,i}$. Then T is a basis for M and the matrix of f w.r.t. T is $B = C^{-1}AC$. In other words, conjugation of matrices corresponds to change of basis.

Proof The proof follows by seeing that the following diagram is commutative.



The diagram also explains what it means for A to be the matrix of f w.r.t. the basis S. Let $h : R^n \to M$ be the isomorphism with $h(e_i) = u_i$ for $1 \leq i \leq n$. Then the matrix $A \in R_n$ is the one determined by the endomorphism $h^{-1} \circ f \circ h : R^n \to R^n$. In other words, column i of A is $h^{-1}(f(h(e_i)))$.

An important special case is where $M = R^n$ and $f : R^n \to R^n$ is given by some matrix W. Then h is given by the matrix U whose i^{th} column is u_i and $A = U^{-1}WU$. In other words, W represents f w.r.t. the standard basis, and $U^{-1}WU$ represents f w.r.t. the basis $\{u_1, .., u_n\}$.

Definition Suppose M is a f.g. free module and $f : M \to M$ is a homomorphism. Define $|f|$ to be $|A|$, trace(f) to be trace(A), and $CP_f(x)$ to be $CP_A(x)$, where A

is the matrix of f w.r.t. some basis. By the previous theorem, all three are well defined, i.e., do not depend upon the choice of basis.

Exercise Let $R = \mathbf{Z}$ and $f : \mathbf{Z}^2 \to \mathbf{Z}^2$ be defined by $f(D) = \begin{pmatrix} 3 & 3 \\ 0 & -1 \end{pmatrix} D$. Find the matrix of f w.r.t. the basis $\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$.

Exercise Let $L \subset \mathbf{R}^2$ be the line $L = \{(r, 2r)^t : r \in \mathbf{R}\}$. Show there is one and only one homomorphism $f : \mathbf{R}^2 \to \mathbf{R}^2$ which is the identity on L and has $f((-1,1)^t) = ((1,-1)^t)$. Find the matrix $A \in \mathbf{R}_2$ which represents f with respect to the basis $\{(1,2)^t, (-1,1)^t\}$. Find the determinant, trace, and characteristic polynomial of f. Also find the matrix $B \in \mathbf{R}_2$ which represents f with respect to the standard basis. Finally, find an invertible matrix $C \in \mathbf{R}_2$ with $B = C^{-1}AC$.

Vector Spaces

So far in this chapter we have been developing the theory of linear algebra in general. The previous theorem, for example, holds for any commutative ring R, but it must be assumed that the module M is free. Endomorphisms in general will not have a determinant, trace, or characteristic polynomial. We now focus on the case where R is a field F, and show that in this case, every F-module is free. Thus any finitely generated F-module will have a well defined dimension, and endomorphisms on it will have well defined determinant, trace, and characteristic polynomial.

In this section, F is a field. F-modules may also be called *vector spaces* and F-module homomorphisms may also be called *linear transformations*.

Theorem Suppose M is an F-module and $v \in M$. Then $v \neq \underline{0}$ iff v is independent. That is, if $v \in V$ and $r \in F$, $vr = \underline{0}$ implies $v = \underline{0}$ in M or $r = \underline{0}$ in F.

Proof Suppose $vr = \underline{0}$ and $r \neq \underline{0}$. Then $\underline{0} = (vr)r^{-1} = v\underline{1} = v$.

Theorem Suppose $M \neq \underline{0}$ is an F-module and $v \in M$. Then v generates M iff v is a basis for M. Furthermore, if these conditions hold, then $M \approx F_F$, any non-zero element of M is a basis, and any two elements of M are dependent.

Proof Suppose v generates M. Then $v \neq \underline{0}$ and is thus independent by the previous theorem. In this case $M \approx F$, and any non-zero element of F is a basis, and any two elements of F are dependent.

Theorem Suppose $M \neq \underline{0}$ is a finitely generated F-module. If $S = \{v_1, .., v_m\}$ generates M, then any maximal independent subsequence of S is a basis for M. Thus any finite independent sequence can be extended to a basis. In particular, M has a finite free basis, and thus is a free F-module.

Proof Suppose, for notational convenience, that $\{v_1, .., v_n\}$ is a maximal independent subsequence of S, and $n < i \leq m$. It must be shown that v_i is a linear combination of $\{v_1, .., v_n\}$. Since $\{v_1, .., v_n, v_i\}$ is dependent, $\exists\ r_1, ..., r_n, r_i$ not all zero, such that $v_1r_1 + \cdots + v_nr_n + v_ir_i = \underline{0}$. Then $r_i \neq \underline{0}$ and $v_i = -(v_1r_1 + \cdots + v_nr_n)r_i^{-1}$. Thus $\{v_1, .., v_n\}$ generates S and thus all of M. Now suppose T is a finite independent sequence. T may be extended to a finite generating sequence, and inside that sequence it may be extended to a maximal independent sequence. Thus T extends to a basis.

After so many routine theorems, it is nice to have one with real power. It not only says any finite independent sequence can be extended to a basis, but it can be extended to a basis inside any finite generating set containing it. This is one of the theorems that makes linear algebra tick. The key hypothesis here is that the ring is a field. If $R = \mathbf{Z}$, then $\mathbf{Z}$ is a free module over itself, and the element 2 of $\mathbf{Z}$ is independent. However it certainly cannot be extended to a basis. Also the finiteness hypothesis in this theorem is only for convenience, as will be seen momentarily.

Since F is a commutative ring, any two bases of M must have the same number of elements, and thus the dimension of M is well defined (see theorem on page 83).

Theorem Suppose M is an F-module of dimension n, and $\{v_1, ..., v_m\}$ is an independent sequence in M. Then $m \leq n$ and if $m = n$, $\{v_1, .., v_m\}$ is a basis.

Proof $\{v_1, .., v_m\}$ extends to a basis with n elements.

The next theorem is just a collection of observations.

Theorem Suppose M and N are finitely generated F-modules.

1) $M \approx F^n$ iff $\dim(M) = n$.
2) $M \approx N$ iff $\dim(M) = \dim(N)$.
3) $F^m \approx F^n$ iff $n = m$.
4) $\dim(M \oplus N) = \dim(M) + \dim(N)$.

Here is the basic theorem for vector spaces in full generality.

Theorem Suppose $M \neq \underline{0}$ is an F-module and $S = \{v_t\}_{t \in T}$ generates M.

1) Any maximal independent subsequence of S is a basis for M.
2) Any independent subsequence of S may be extended to a maximal independent subsequence of S, and thus to a basis for M.
3) Any independent subsequence of M can be extended to a basis for M. In particular, M has a free basis, and thus is a free F-module.

Proof The proof of 1) is the same as in the case where S is finite. Part 2) will follow from the Hausdorff Maximality Principle. An independent subsequence of S is contained in a maximal monotonic tower of independent subsequences. The union of these independent subsequences is still independent, and so the result follows. Part 3) follows from 2) because an independent sequence can always be extended to a generating sequence.

Theorem Suppose M is an F-module and $K \subset M$ is a submodule.

1) K is a summand of M, i.e., $\exists$ a submodule L of M with $K \oplus L = M$.
2) If M is f.g., then $\dim(K) \leq \dim(M)$ and $K = M$ iff $\dim(K) = \dim(M)$.

Proof Let T be a basis for K. Extend T to a basis S for M. Then $S - T$ generates a submodule L with $K \oplus L = M$. Part 2) follows from 1).

Corollary $\mathbf{Q}$ is a summand of $\mathbf{R}$. In other words, $\exists$ a $\mathbf{Q}$-submodule $V \subset \mathbf{R}$ with $\mathbf{Q} \oplus V = \mathbf{R}$ as $\mathbf{Q}$-modules. (See exercise on page 77.)

Proof $\mathbf{Q}$ is a field, $\mathbf{R}$ is a $\mathbf{Q}$-module, and $\mathbf{Q}$ is a submodule of $\mathbf{R}$.

Corollary Suppose M is a f.g. F-module, N is an F-module, and $f : M \to N$ is a homomorphism. Then $\dim(M) = \dim(\ker(f)) + \dim(\text{image}(f))$.

Proof Let $K = \ker(f)$ and $L \subset M$ be a submodule with $K \oplus L = M$. Then $f \mid L : L \to \text{image}(f)$ is an isomorphism.

Exercise Suppose R is a domain with the property that, for R-modules, every submodule is a summand. Show R is a field.

Exercise Find a free $\mathbf{Z}$-module which has a generating set containing no basis.

Exercise The real vector space $\mathbf{R}^2$ is generated by the sequence $S = \{(\pi, 0), (2, 1), (3, 2)\}$. Show there are three maximal independent subsequences of S, and each is a basis for $\mathbf{R}^2$. (Row vectors are used here just for convenience.)

The real vector space $\mathbf{R}^3$ is generated by $S = \{(1, 1, 2), (1, 2, 1), (3, 4, 5), (1, 2, 0)\}$. Show there are three maximal independent subsequences of S and each is a basis for $\mathbf{R}^3$. You may use determinant.

Square matrices over fields

This theorem is just a summary of what we have for square matrices over fields.

Theorem Suppose $A \in F_n$ and $f : F^n \to F^n$ is defined by $f(B) = AB$. Let $v_1, .., v_n \in F^n$ be the columns of A, and $w_1, .., w_n \in F^n = F_{1,n}$ be the rows of A. Then the following are equivalent

1) $\{v_1, .., v_n\}$ is independent, i.e., f is injective.

2) $\{v_1, .., v_n\}$ is a basis for F^n, i.e., f is an automorphism, i.e., A is invertible, i.e., $\mid A \mid \neq \underline{0}$.

3) $\{v_1, .., v_n\}$ generates F^n, i.e., f is surjective.

1^t) $\{w_1, .., w_n\}$ is independent.

2^t) $\{w_1, .., w_n\}$ is a basis for F^n, i.e., A^t is invertible, i.e., $\mid A^t \mid \neq \underline{0}$.

3^t) $\{w_1, .., w_n\}$ generates F^n.

Proof Except for 1) and 1^t), this theorem holds for any commutative ring R. (See the section **Relating these concepts to square matrices**, pages 81 and 82.) Parts 1) and 1^t) follow from the preceding section.

Exercise Add to this theorem more equivalent statements in terms of solutions of n equations in n unknowns.

Overview Suppose each of X and Y is a set with n elements and $f : X \to Y$ is a function. By the pigeonhole principle, f is injective iff f is bijective iff f is surjective. Now suppose each of U and V is a vector space of dimension n and $f : U \to V$ is a linear transformation. It follows from the work done so far that f is injective iff f is bijective iff f is surjective. This shows some of the simple and definitive nature of linear algebra.

Exercise Let $A = (A_1, .., A_n)$ be an $n \times n$ matrix over $\mathbf{Z}$ with column $i = A_i \in \mathbf{Z}^n$. Let $f : \mathbf{Z}^n \to \mathbf{Z}^n$ be defined by $f(B) = AB$ and $\bar{f} : \mathbf{R}^n \to \mathbf{R}^n$ be defined by $\bar{f}(C) = AC$. Show the following are equivalent. (See the exercise on page 79.)

1) $f : \mathbf{Z}^n \to \mathbf{Z}^n$ is injective.

2) The sequence $(A_1, .., A_n)$ is linearly independent over $\mathbf{Z}$.

3) $|A| \neq 0$.

4) $\bar{f} : \mathbf{R}^n \to \mathbf{R}^n$ is injective.

5) The sequence $(A_1, .., A_n)$ is linearly independent over $\mathbf{R}$.

Rank of a matrix Suppose $A \in F_{m,n}$. The row (column) rank of A is defined to be the dimension of the submodule of F^n (F^m) generated by the rows (columns) of A.

Theorem If $C \in F_m$ and $D \in F_n$ are invertible, then the row (column) rank of A is the same as the row (column) rank of CAD.

Proof Suppose $f : F^n \to F^m$ is defined by $f(B) = AB$. Each column of A is a vector in the range F^m, and we know from page 81 that each $f(B)$ is a linear

combination of those vectors. Thus the image of f is the submodule of F^m generated by the columns of A, and its dimension is the column rank of A. This dimension is the same as the dimension of the image of $g \circ f \circ h : F^n \to F^m$, where h is any automorphism on F^n and g is any automorphism on F^m. This proves the theorem for column rank. The theorem for row rank follows using transpose.

Theorem If $A \in F_{m,n}$, the row rank and the column rank of A are equal. This number is called the *rank* of A and is $\leq \min\{m, n\}$.

Proof By the theorem above, elementary row and column operations change neither the row rank nor the column rank. By row and column operations, A may be changed to a matrix H where $h_{1,1} = \cdot\cdot = h_{t,t} = \underline{1}$ and all other entries are $\underline{0}$ (see the first exercise on page 59). Thus row rank $= t =$ column rank.

Exercise Suppose A has rank t. Show that it is possible to select t rows and t columns of A such that the determined $t \times t$ matrix is invertible. Show that the rank of A is the largest integer t such that this is possible.

Exercise Suppose $A \in F_{m,n}$ has rank t. What is the dimension of the solution set of $AX = \underline{0}$?

Definition Suppose M is a finite dimensional vector space over a field F, and $f : M \to M$ is an endomorphism. The *rank* of f is defined to be the dimension of the image of f. It follows from the work above that this is the same as the rank of any matrix representing f.

Geometric Interpretation of Determinant

Suppose $V \subset \mathbf{R}^n$ is some nice subset. For example, if $n = 2$, V might be the interior of a square or circle. There is a concept of the n-dimensional volume of V. For $n = 1$, it is length. For $n = 2$, it is area, and for $n = 3$ it is "ordinary volume". Suppose $A \in \mathbf{R}_n$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is the homomorphism given by A. The volume of V does not change under translation, i.e., V and $V + p$ have the same volume. Thus $f(V)$ and $f(V + p) = f(V) + f(p)$ have the same volume. In street language, the next theorem says that "f multiplies volume by the absolute value of its determinant".

Theorem The n-dimensional volume of $f(V)$ is $\pm|A|$(the n-dimensional volume of V). Thus if $|A| = \pm 1$, f preserves volume.

Proof If $|A| = 0$, image(f) has dimension $< n$ and thus $f(V)$ has n-dimensional volume 0. If $|A| \neq 0$ then A is the product of elementary matrices (see page 59) and for elementary matrices, the theorem is obvious. The result follows because the determinant of the composition is the product of the determinants.

Corollary If P is the n-dimensional parallelepiped determined by the columns $v_1, .., v_n$ of A, then the n-dimensional volume of P is $\pm|A|$.

Proof Let $V = [0,1] \times \cdots \times [0,1] = \{e_1t_1 + \cdots + e_nt_n : 0 \leq t_i \leq 1\}$. Then $P = f(V) = \{v_1t_1 + \cdots + v_nt_n : 0 \leq t_i \leq 1\}$.

Linear functions approximate differentiable functions locally

We continue with the special case $F = \mathbf{R}$. Linear functions arise naturally in business, science, and mathematics. However this is not the only reason that linear algebra is so useful. It is a central fact that smooth phenomena may be approximated locally by linear phenomena. Without this great simplification, the world of technology as we know it today would not exist. Of course, linear transformations send the origin to the origin, so they must be adjusted by a translation. As a simple example, suppose $h : \mathbf{R} \to \mathbf{R}$ is differentiable and p is a real number. Let $f : \mathbf{R} \to \mathbf{R}$ be the linear transformation $f(x) = h'(p)x$. Then h is approximated near p by $g(x) = h(p) + f(x - p) = h(p) + h'(p)(x - p)$.

Now suppose $V \subset \mathrm{R}^2$ is some nice subset and $h = (h_1, h_2) : V \to \mathbf{R}^2$ is injective and differentiable. Define the Jacobian by $J(h)(x,y) = \begin{pmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{pmatrix}$ and for each $(x,y) \in V$, let $f(x,y) : \mathbf{R}^2 \to \mathbf{R}^2$ be the homomorphism defined by $J(h)(x,y)$. Then for any $(p_1, p_2) \in V$, h is approximated near (p_1, p_2) (after translation) by $f(p_1, p_2)$. The area of V is $\int\int_V 1 dxdy$. From the previous section we know that any homomorphism f multiplies area by $\mid f \mid$. The student may now understand the following theorem from calculus. (Note that if h is the restriction of a linear transformation from $\mathbf{R}^2$ to $\mathbf{R}^2$, this theorem is immediate from the previous section.)

Theorem Suppose the determinant of $J(h)(x,y)$ is non-negative for each $(x,y) \in V$. Then the area of $h(V)$ is $\int\int_V \mid J(h) \mid dxdy$.

The Transpose Principle

We now return to the case where F is a field (of arbitrary characteristic). F-modules may also be called *vector spaces* and submodules may be called *subspaces*. The study of R-modules in general is important and complex. However the study of F-modules is short and simple – every vector space is free and every subspace is a summand. The core of classical linear algebra is not the study of vector spaces, but the study of homomorphisms, and in particular, of endomorphisms. One goal is to show that if $f : V \to V$ is a homomorphism with some given property, there exists a basis of V so that the matrix representing f displays that property in a prominent manner. The next theorem is an illustration of this.

Theorem Let F be a field and n be a positive integer.

1) Suppose V is an n-dimensional vector space and $f : V \to V$ is a homomorphism with $|f| = \underline{0}$. Then $\exists$ a basis of V such that the matrix representing f has its first row zero.
2) Suppose $A \in F_n$ has $|A| = \underline{0}$. Then $\exists$ an invertible matrix C such that $C^{-1}AC$ has its first row zero.
3) Suppose V is an n-dimensional vector space and $f : V \to V$ is a homomorphism with $|f| = 0$. Then $\exists$ a basis of V such that the matrix representing f has its first column zero.
4) Suppose $A \in F_n$ has $|A| = \underline{0}$. Then $\exists$ an invertible matrix D such that $D^{-1}AD$ has its first column zero.

We first wish to show that these 4 statements are equivalent. We know that 1) and 2) are equivalent and also that 3) and 4) are equivalent because change of basis corresponds to conjugation of the matrix. Now suppose 2) is true and show 4) is true. Suppose $|A| = \underline{0}$. Then $|A^t| = \underline{0}$ and by 2) $\exists$ C such that $C^{-1}A^tC$ has first row zero. Thus $(C^{-1}A^tC)^t = C^tA(C^t)^{-1}$ has first row column zero. The result follows by defining $D = (C^t)^{-1}$. Also 4) implies 2).

This is an example of the *transpose principle*. Loosely stated, it is that theorems about change of basis correspond to theorems about conjugation of matrices and theorems about the rows of a matrix correspond to theorems about the columns of a matrix, using transpose. In the remainder of this chapter, this will be used without further comment.

Proof of the theorem We are free to select any of the 4 parts, and we select part 3). Since $| f | = 0$, f is not injective and $\exists$ a non-zero $v_1 \in V$ with $f(v_1) = \underline{0}$. Now v_1 is independent and extends to a basis $\{v_1, .., v_n\}$. Then the matrix of f w.r.t this basis has first column zero.

Exercise Let $A = \begin{pmatrix} 3\pi & 6 \\ 2\pi & 4 \end{pmatrix}$. Find an invertible matrix $C \in \mathbf{R}_2$ so that $C^{-1}AC$ has first row zero. Also let $A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 3 & 4 \\ 2 & 1 & 4 \end{pmatrix}$ and find an invertible matrix $D \in \mathbf{R}_3$ so that $D^{-1}AD$ has first column zero.

Exercise Suppose M is an n-dimensional vector space over a field F, k is an integer with $0 < k < n$, and $f : M \to M$ is an endomorphism of rank k. Show there is a basis for M so that the matrix representing f has its first $n - k$ rows zero. Also show there is a basis for M so that the matrix representing f has its first $n - k$ columns zero. Work this out without using the transpose principle.

Nilpotent Homomorphisms

In this section it is shown that an endomorphism f is nilpotent iff all of its characteristic roots are $\underline{0}$ iff it may be represented by a strictly upper triangular matrix.

Definition An endomorphism $f : V \to V$ is nilpotent if $\exists \ \ m$ with $f^m = \underline{0}$. Any f represented by a strictly upper triangular matrix is nilpotent (see page 56).

Theorem Suppose V is an n-dimensional vector space and $f : V \to V$ is a nilpotent homomorphism. Then $f^n = \underline{0}$ and $\exists$ a basis of V such that the matrix representing f w.r.t. this basis is strictly upper triangular. Thus the characteristic polynomial of f is $CP_f(x) = x^n$.

Proof Suppose $f \neq \underline{0}$ is nilpotent. Let t be the largest positive integer with $f^t \neq \underline{0}$. Then $f^t(V) \subset f^{t-1}(V) \subset \cdot\cdot \subset f(V) \subset V$. Since f is nilpotent, all of these inclusions are proper. Therefore $t < n$ and $f^n = \underline{0}$. Construct a basis for V by starting with a basis for $f^t(V)$, extending it to a basis for $f^{t-1}(V)$, etc. Then the matrix of f w.r.t. this basis is strictly upper triangular.

Note To obtain a matrix which is strictly lower triangular, reverse the order of the basis.

Exercise Use the transpose principle to write 3 other versions of this theorem.

Theorem Suppose V is an n-dimensional vector space and $f : V \to V$ is a homomorphism. Then f is nilpotent iff $CP_f(x) = x^n$. (See the exercise at the end of Chapter 4 for the case $n = 2$.)

Proof Suppose $CP_f(x) = x^n$. For $n = 1$ this implies $f = \underline{0}$, so suppose $n > 1$. Since the constant term of $CP_f(x)$ is $\underline{0}$, the determinant of f is $\underline{0}$. Thus $\exists$ a basis of V such that the matrix A representing f has its first column zero. Let $B \in F_{n-1}$ be the matrix obtained from A by removing its first row and first column. Now $CP_A(x) = x^n = xCP_B(x)$. Thus $CP_B(x) = x^{n-1}$ and by induction on n, B is nilpotent and so $\exists$ C such that $C^{-1}BC$ is strictly upper triangular. Then

$$\begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & & & & \\ \cdot & & C^{-1} & & \\ \cdot & & & & \\ 0 & & & & \end{pmatrix} \begin{pmatrix} 0 & * & \cdot & \cdot & * \\ \cdot & & & & \\ & & B & & \\ \cdot & & & & \\ 0 & & & & \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & & & & \\ \cdot & & C & & \\ \cdot & & & & \\ 0 & & & & \end{pmatrix} = \begin{pmatrix} 0 & * & \cdot & \cdot & * \\ 0 & & & & \\ \cdot & & C^{-1}BC & & \\ \cdot & & & & \\ 0 & & & & \end{pmatrix}$$

is strictly upper triangular.

Exercise Suppose F is a field, $A \in F_3$ is a lower triangular matrix of rank 2, and $B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Using conjugation by elementary matrices, show there is an invertible matrix C so that $C^{-1}AC = B$. Now suppose V is a 3-dimensional vector space and $f : V \to V$ is a nilpotent endomorphism of rank 2. We know f can be represented by a lower triangular matrix. Show there is a basis $\{v_1, v_2, v_3\}$ for V so that B is the matrix representing f. Also show that $f(v_1) = v_2$, $f(v_2) = v_3$, and $f(v_3) = \underline{0}$. In other words, there is a basis for V of the form $\{v, f(v), f^2(v)\}$ with $f^3(v) = \underline{0}$.

Exercise Suppose V is a 3-dimensional vector space and $f : V \to V$ is a nilpotent endomorphism of rank 1. Show there is a basis for V so that the matrix representing f is $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Eigenvalues

Our standing hypothesis is that V is an n-dimensional vector space over a field F and $f : V \to V$ is a homomorphism.

Definition An element $\lambda \in F$ is an *eigenvalue* of f if $\exists$ a non-zero $v \in V$ with $f(v) = \lambda v$. Any such v is called an *eigenvector*. $E_\lambda \subset V$ is defined to be the set of all eigenvectors for λ (plus $\underline{0}$). Note that $E_\lambda = \ker(\lambda I - f)$ is a subspace of V. The next theorem shows the eigenvalues of f are just the characteristic roots of f.

Theorem If $\lambda \in F$ then the following are equivalent.

1) λ is an eigenvalue of f, i.e., $(\lambda I - f) : V \to V$ is not injective.
2) $\mid (\lambda I - f) \mid = \underline{0}$.
3) λ is a characteristic root of f, i.e., a root of the characteristic polynomial $CP_f(x) = \mid (xI - A) \mid$, where A is any matrix representing f.

Proof It is immediate that 1) and 2) are equivalent, so let's show 2) and 3) are equivalent. The evaluation map $F[x] \to F$ which sends $h(x)$ to $h(\lambda)$ is a ring homomorphism (see theorem on page 47). So evaluating $(xI - A)$ at $x = \lambda$ and taking determinant gives the same result as taking the determinant of $(xI - A)$ and evaluating at $x = \lambda$. Thus 2) and 3) are equivalent.

The nicest thing you can say about a matrix is that it is similar to a diagonal matrix. Here is one case where that happens.

Theorem Suppose $\lambda_1, .., \lambda_k$ are distinct eigenvalues of f, and v_i is an eigenvector of λ_i for $1 \leq i \leq k$. Then the following hold.

1) $\{v_1, .., v_k\}$ is independent.

2) If $k = n$, i.e., if $CP_f(x) = (x - \lambda_1) \cdots (x - \lambda_n)$, then $\{v_1, .., v_n\}$ is a basis for V. The matrix of f w.r.t. this basis is the diagonal matrix whose (i, i) term is λ_i.

Proof Suppose $\{v_1, .., v_k\}$ is dependent. Suppose t is the smallest positive integer such that $\{v_1, .., v_t\}$ is dependent, and $v_1 r_1 + \cdots + v_t r_t = \underline{0}$ is a non-trivial linear combination. Note that at least two of the coefficients must be non-zero. Now $(f - \lambda_t)(v_1 r_1 + \cdots + v_t r_t) = v_1(\lambda_1 - \lambda_t) r_1 + \cdots + v_{t-1}(\lambda_{t-1} - \lambda_t) r_{t-1} + \underline{0} = \underline{0}$ is a shorter

non-trivial linear combination. This is a contradiction and proves 1). Part 2) follows from 1) because $\dim(V) = n$.

Exercise Let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbf{R}_2$. Find an invertible $C \in \mathbf{C}_2$ such that $C^{-1}AC$ is diagonal. Show that C cannot be selected in $\mathbf{R}_2$. Find the characteristic polynomial of A.

Exercise Suppose V is a 3-dimensional vector space and $f : V \to V$ is an endomorphism with $CP_f(x) = (x-\lambda)^3$. Show that $(f - \lambda I)$ has characteristic polynomial x^3 and is thus a nilpotent endomorphism. Show there is a basis for V so that the matrix representing f is $\begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{pmatrix}$, $\begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$ or $\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$.

We could continue and finally give an ad hoc proof of the Jordan canonical form, but in this chapter we prefer to press on to inner product spaces. The Jordan form will be developed in Chapter 6 as part of the general theory of finitely generated modules over Euclidean domains. The next section is included only as a convenient reference.

Jordan Canonical Form

This section should be just skimmed or omitted entirely. It is unnecessary for the rest of this chapter, and is not properly part of the flow of the chapter. The basic facts of Jordan form are summarized here simply for reference.

The statement that a square matrix B over a field F is a *Jordan block* means that $\exists\ \lambda \in F$ such that B is a lower triangular matrix of the form

$$B = \begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \cdot & & \\ & & \cdot & \\ 0 & & 1 & \lambda \end{pmatrix}.$$ B gives a homomorphism $g : F^m \to F^m$ with $g(e_m) = \lambda e_m$

and $g(e_i) = e_{i+1} + \lambda e_i$ for $1 \leq i < m$. Note that $CP_B(x) = (x-\lambda)^m$ and so λ is the only eigenvalue of B, and B satisfies its characteristic polynomial, i.e., $CP_B(B) = \underline{0}$.

Definition A matrix $D \in F_n$ is in Jordan form if $\exists$ Jordan blocks $B_1, .., B_t$ such that $D = \begin{pmatrix} B_1 & & & \\ & B_2 & & 0 \\ & & \ddots & \\ & 0 & & B_t \end{pmatrix}$. Suppose D is of this form and $B_i \in F_{n_i}$ has eigenvalue λ_i. Then $n_1 + \cdots + n_t = n$ and $CP_D(x) = (x - \lambda_1)^{n_1} \cdot\cdot (x - \lambda_t)^{n_t}$. Note that a diagonal matrix is a special case of Jordan form. D is a diagonal matrix iff each $n_i = 1$, i.e., iff each Jordan block is a 1×1 matrix.

Theorem If $A \in F_n$, the following are equivalent.

1) $\exists$ an invertible $C \in F_n$ such that $C^{-1}AC$ is in Jordan form.

2) $\exists\ \lambda_1, .., \lambda_n \in F$ (not necessarily distinct) such that $CP_A(x) = (x - \lambda_1) \cdot\cdot (x - \lambda_n)$. (In this case we say that all the eigenvalues of A belong to F.)

Theorem Jordan form (when it exists) is unique. This means that if A and D are similar matrices in Jordan form, they have the same Jordan blocks, except possibly in different order.

The reader should use the transpose principle to write three other versions of the first theorem. Also note that we know one special case of this theorem, namely that if A has n distinct eigenvalues in F, then A is similar to a diagonal matrix. Later on it will be shown that if A is a symmetric real matrix, then A is similar to a diagonal matrix.

Let's look at the classical case $A \in \mathbf{R}_n$. The complex numbers are algebraically closed. This means that $CP_A(x)$ will factor completely in $\mathbf{C}[x]$, and thus $\exists\ C \in \mathbf{C}_n$ with $C^{-1}AC$ in Jordan form. C may be selected to be in $\mathbf{R}_n$ iff all the eigenvalues of A are real.

Exercise Find all real matrices in Jordan form that have the following characteristic polynomials: $x(x-2)$, $(x-2)^2$, $(x-2)(x-3)(x-4)$, $(x-2)(x-3)^2$, $(x-2)^2(x-3)^2$, $(x-2)(x-3)^3$.

Exercise Suppose $D \in F_n$ is in Jordan form and has characteristic polynomial $a_0 + a_1x + \cdots + x^n$. Show $a_0I + a_1D + \cdots + D^n = \underline{0}$, i.e., show $CP_D(D) = \underline{0}$.

Exercise (Cayley-Hamilton Theorem) Suppose E is a field and $A \in E_n$. Assume the theorem that there is a field F containing E such that $CP_A(x)$ factors completely in $F[x]$. Thus $\exists$ an invertible $C \in F_n$ such that $D = C^{-1}AC$ is in Jordan form. Use this to show $CP_A(A) = \underline{0}$. (See the second exercise on page 66.)

Exercise Suppose $A \in F_n$ is in Jordan form. Show A is nilpotent iff $A^n = \underline{0}$ iff $CP_A(x) = x^n$. (Note how easy this is in Jordan form.)

Inner Product Spaces

The two most important fields for mathematics and science in general are the real numbers and the complex numbers. Finitely generated vector spaces over $\mathbf{R}$ or $\mathbf{C}$ support inner products and are thus geometric as well as algebraic objects. The theories for the real and complex cases are quite similar, and both could have been treated here. However, for simplicity, attention is restricted to the case $F = \mathbf{R}$. In the remainder of this chapter, the power and elegance of linear algebra become transparent for all to see.

Definition Suppose V is a real vector space. An *inner product* (or *dot product*) on V is a function $V \times V \to \mathbf{R}$ which sends (u, v) to $u \cdot v$ and satisfies

1) $(u_1r_1 + u_2r_2) \cdot v = (u_1 \cdot v)r_1 + (u_2 \cdot v)r_2$
$v \cdot (u_1r_1 + u_2r_2) = (v \cdot u_1)r_1 + (v \cdot u_2)r_2$
for all $u_1, u_2, v \in V$ and $r_1, r_2 \in \mathbf{R}$.

2) $u \cdot v = v \cdot u$ for all $u, v \in V$.

3) $u \cdot u \geq 0$ and $u \cdot u = 0$ iff $u = \underline{0}$ for all $u \in V$.

Theorem Suppose V has an inner product.

1) If $v \in V$, $f : V \to \mathbf{R}$ defined by $f(u) = u \cdot v$ is a homomorphism. Thus $\underline{0} \cdot v = 0$.

2) Schwarz' inequality. If $u, v \in V$, $(u \cdot v)^2 \leq (u \cdot u)(v \cdot v)$.

Proof of 2) Let $a = \sqrt{v \cdot v}$ and $b = \sqrt{u \cdot u}$. If a or b is 0, the result is obvious. Suppose neither a nor b is 0. Now $0 \leq (ua \pm vb) \cdot (ua \pm vb) = (u \cdot u)a^2 \pm 2ab(u \cdot v) + (v \cdot v)b^2 = b^2a^2 \pm 2ab(u \cdot v) + a^2b^2$. Dividing by $2ab$ yields $0 \leq ab \pm (u \cdot v)$ or $| u \cdot v | \leq ab$.

Theorem Suppose V has an inner product. Define the *norm* or *length* of a vector v by $\|v\| = \sqrt{v \cdot v}$. The following properties hold.

1) $\|v\| = 0$ iff $v = \underline{0}$.

2) $\|vr\| = \|v\| \mid r \mid$.

3) $\mid u \cdot v \mid \leq \|u\|\|v\|$. (Schwarz' inequality)

4) $\|u + v\| \leq \|u\| + \|v\|$. (The triangle inequality)

Proof of 4) $\|u + v\|^2 = (u + v) \cdot (u + v) = \|u\|^2 + 2(u \cdot v) + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$.

Definition An Inner Product Space (IPS) is a real vector space with an inner product. Suppose V is an IPS. A sequence $\{v_1, .., v_n\}$ is *orthogonal* provided $v_i \cdot v_j = 0$ when $i \neq j$. The sequence is *orthonormal* if it is orthogonal and each vector has length 1, i.e., $v_i \cdot v_j = \delta_{i,j}$ for $1 \leq i, j \leq n$.

Theorem If $S = \{v_1, .., v_n\}$ is an orthogonal sequence of non-zero vectors in an IPS V, then S is independent. Furthermore $\left\{ \dfrac{v_1}{\|v_1\|}, \cdots, \dfrac{v_n}{\|v_n\|} \right\}$ is orthonormal.

Proof Suppose $v_1 r_1 + \cdot\cdot + v_n r_n = \underline{0}$. Then $0 = (v_1 r_1 + \cdot\cdot + v_n r_n) \cdot v_i = r_i (v_i \cdot v_i)$ and thus $r_i = 0$. Thus S is independent. The second statement is transparent.

It is easy to define an inner product, as is shown by the following theorem.

Theorem Suppose V is a real vector space with a basis $S = \{v_1, .., v_n\}$. Then there is a unique inner product on V which makes S an orthornormal basis. It is given by the formula $(v_1 r_1 + \cdot\cdot + v_n r_n) \cdot (v_1 s_1 + \cdot\cdot + v_n s_n) = r_1 s_1 + \cdot\cdot + r_n s_n$.

Convention $\mathbf{R}^n$ will be assumed to have the *standard inner product* defined by $(r_1, .., r_n)^t \cdot (s_1, .., s_n)^t = r_1 s_1 + \cdot\cdot + r_n s_n$. $S = \{e_1, .., e_n\}$ will be called the *canonical* or *standard orthonormal basis* (see page 72). The next theorem shows that this inner product has an amazing geometry.

Theorem If $u, v \in \mathbf{R}^n$, $u \cdot v = \|u\|\|v\| \cos \Theta$ where Θ is the angle between u

and v.

Proof Let $u = (r_1, .., r_n)$ and $v = (s_1, .., s_n)$. By the law of cosines $\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \cos \Theta$. So $(r_1 - s_1)^2 + \cdots + (r_n - s_n)^2 = r_1^2 + \cdots + r_n^2 + s_1^2 + \cdots + s_n^2 - 2\|u\|\|v\| \cos \Theta$. Thus $r_1 s_1 + \cdots + r_n s_n = \|u\|\|v\| \cos \Theta$.

Exercise This is a simple exercise to observe that hyperplanes in $\mathbf{R}^n$ are cosets. Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is a non-zero homomorphism given by a matrix $A = (a_1, .., a_n) \in \mathbf{R}_{1,n}$. Then $L = \ker(f)$ is the set of all solutions to $a_1 x_1 + \cdots + a_n x_n = 0$, i.e., the set of all vectors perpendicular to A. Now suppose $b \in \mathbf{R}$ and $C = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_n \end{pmatrix} \in \mathbf{R}^n$ has $f(C) = b$. Then $f^{-1}(b)$ is the set of all solutions to $a_1 x_1 + \cdots + a_n x_n = b$ which is the coset $L + C$, and this the set of all solutions to $a_1(x_1 - c_1) + \cdots + a_n(x_n - c_n) = 0$.

Gram-Schmidt orthonormalization

Theorem (Fourier series) Suppose W is an IPS with an orthonormal basis $\{w_1, .., w_n\}$. Then if $v \in W$, $v = w_1(v \cdot w_1) + \cdots + w_n(v \cdot w_n)$.

Proof $v = w_1 r_1 + \cdots + w_n r_n$ and $v \cdot w_i = (w_1 r_1 + \cdots + w_n r_n) \cdot w_i = r_i$

Theorem Suppose W is an IPS, $Y \subset W$ is a subspace with an orthonormal basis $\{w_1, .., w_k\}$, and $v \in W - Y$. Define the *projection* of v onto Y by $p(v) = w_1(v \cdot w_1) + \cdots + w_k(v \cdot w_k)$, and let $w = v - p(v)$. Then $(w \cdot w_i) = (v - w_1(v \cdot w_1) \cdots - w_k(v \cdot w_k)) \cdot w_i = 0$. Thus if $w_{k+1} = \dfrac{w}{\|w\|}$, then $\{w_1, .., w_{k+1}\}$ is an orthonormal basis for the subspace generated by $\{w_1, .., w_k, v\}$. If $\{w_1, .., w_k, v\}$ is already orthonormal, $w_{k+1} = v$.

Theorem (Gram-Schmidt) Suppose W is an IPS with a basis $\{v_1, .., v_n\}$. Then W has an orthonormal basis $\{w_1, .., w_n\}$. Moreover, any orthonormal sequence in W extends to an orthonormal basis of W.

Proof Let $w_1 = \dfrac{v_1}{\|v_1\|}$. Suppose inductively that $\{w_1, .., w_k\}$ is an orthonormal basis for Y, the subspace generated by $\{v_1, .., v_k\}$. Let $w = v_{k+1} - p(v_{k+1})$ and

$w_{k+1} = \dfrac{w}{\|w\|}$. Then by the previous theorem, $\{w_1, .., w_{k+1}\}$ is an orthonormal basis for the subspace generated by $\{w_1, .., w_k, v_{k+1}\}$. In this manner an orthonormal basis for W is constructed. Notice that this construction defines a function h which sends a basis for W to an orthonormal basis for W (see topology exercise on page 103).

Now suppose W has dimension n and $\{w_1, .., w_k\}$ is an orthonormal sequence in W. Since this sequence is independent, it extends to a basis $\{w_1, .., w_k, v_{k+1}, .., v_n\}$. The process above may be used to modify this to an orthonormal basis $\{w_1, .., w_n\}$.

Exercise Let $f : \mathbf{R}^3 \to \mathbf{R}$ be the homomorphism defined by the matrix (2,1,3). Find an orthonormal basis for the kernel of f.. Find the projection of $(e_1 + e_2)$ onto $\ker(f)$. Find the angle between $e_1 + e_2$ and the plane $\ker(f)$.

Exercise Let $W = \mathbf{R}^3$ have the standard inner product and $Y \subset W$ be the subspace generated by $\{w_1, w_2\}$ where $w_1 = (1, 0, 0)^t$ and $w_2 = (0, 1, 0)^t$. W is generated by the sequence $\{w_1, w_2, v\}$ where $v = (1, 2, 3)^t$. As in the first theorem of this section, let $w = v - p(v)$, where $p(v)$ is the projection of v onto Y, and set $w_3 = \dfrac{w}{\|w\|}$. Find w_3 and show that for any t with $0 \leq t \leq 1$, $\{w_1, w_2, (1-t)v + tw_3\}$ is a basis for W. This is a key observation for an exercise on page 103 showing $O(n)$ is a deformation retract of $GL_n(\mathbf{R})$.

Isometries Suppose each of U and V is an IPS. A homomorphism $f : U \to V$ is said to be an *isometry* provided it is an isomorphism and for any u_1, u_2 in U, $(u_1 \cdot u_2)_U = (f(u_1) \cdot f(u_2))_V$.

Theorem Suppose each of U and V is an n-dimensional IPS, $\{u_1, .., u_n\}$ is an orthonormal basis for U, and $f : U \to V$ is a homomorphism. Then f is an isometry iff $\{f(u_1), .., f(u_n)\}$ is an orthonormal sequence in V.

Proof Isometries certainly preserve orthonormal sequences. So suppose $T = \{f(u_1), .., f(u_n)\}$ is an orthonormal sequence in V. Then T is independent and thus T is a basis for V and thus f is an isomorphism (see the second theorem on page 79). It is easy to check that f preserves inner products.

We now come to one of the definitive theorems in linear algebra. It is that, up to isometry, there is only one inner product space for each dimension.

Theorem Suppose each of U and V is an n-dimensional IPS. Then $\exists$ an isometry $f : U \to V$. In particular, U is isometric to $\mathbf{R}^n$ with its standard inner product.

Proof There exist orthonormal bases $\{u_1, .., u_n\}$ for U and $\{v_1, .., v_n\}$ for V. By the first theorem on page 79, there exists a homomorphism $f : U \to V$ with $f(u_i) = v_i$, and by the previous theorem, f is an isometry.

Exercise Let $f : \mathbf{R}^3 \to \mathbf{R}$ be the homomorphism defined by the matrix (2,1,3). Find a linear transformation $h : \mathbf{R}^2 \to \mathbf{R}^3$ which gives an isometry from $\mathbf{R}^2$ to $\ker(f)$.

Orthogonal Matrices

As noted earlier, linear algebra is not so much the study of vector spaces as it is the study of endomorphisms. We now wish to study isometries from $\mathbf{R}^n$ to $\mathbf{R}^n$.

We know from a theorem on page 90 that an endomorphism preserves volume iff its determinant is ± 1. Isometries preserve inner product, and thus preserve angle and distance, and so certainly preserve volume.

Theorem Suppose $A \in \mathbf{R}_n$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is the homomorphism defined by $f(B) = AB$. Then the following are equivalent.

1) The columns of A form an orthonormal basis for $\mathbf{R}^n$, i.e., $A^tA = I$.
2) The rows of A form an orthonormal basis for $\mathbf{R}^n$, i.e., $AA^t = I$.
3) f is an isometry.

Proof A left inverse of a matrix is also a right inverse (see the exercise on page 64). Thus 1) and 2) are equivalent because each of them says A is invertible with $A^{-1} = A^t$. Now $\{e_1, .., e_n\}$ is the canonical orthonormal basis for $\mathbf{R}^n$, and $f(e_i)$ is column i of A. Thus by the previous section, 1) and 3) are equivalent.

Definition If $A \in \mathbf{R}_n$ satisfies these three conditions, A is said to be *orthogonal*. The set of all such A is denoted by $O(n)$, and is called the *orthogonal group*.

Theorem

1) If A is orthogonal, $\mid A \mid = \pm 1$.

2) If A is orthogonal, A^{-1} is orthogonal. If A and C are orthogonal, AC is orthogonal. Thus $O(n)$ is a multiplicative subgroup of $GL_n(\mathbf{R})$.

3) Suppose A is orthogonal and f is defined by $f(B) = AB$. Then f preserves distances and angles. This means that if $u, v \in \mathbf{R}^n$, $\|u - v\| = \|f(u) - f(v)\|$ and the angle between u and v is equal to the angle between $f(u)$ and $f(v)$.

Proof Part 1) follows from $|A|^2 = |A|\,|A^t| = |I| = 1$. Part 2) is immediate, because isometries clearly form a subgroup of the multiplicative group of all automorphisms. For part 3) assume $f : \mathbf{R}^n \to \mathbf{R}^n$ is an isometry. Then $\|u - v\|^2 = (u - v) \cdot (u - v) = f(u - v) \cdot f(u - v) = \|f(u - v)\|^2 = \|f(u) - f(v)\|^2$. The proof that f preserves angles follows from $u \cdot v = \|u\|\|v\|\cos\Theta$.

Exercise Show that if $A \in O(2)$ has $|A| = 1$, then $A = \begin{pmatrix} \cos\Theta & -\sin\Theta \\ \sin\Theta & \cos\Theta \end{pmatrix}$ for some number Θ. (See the exercise on page 56.)

Exercise (topology) Let $\mathbf{R}_n \approx \mathbf{R}^{n^2}$ have its usual metric topology. This means a sequence of matrices $\{A_i\}$ converges to A iff it converges coordinatewise. Show $GL_n(\mathbf{R})$ is an open subset and $O(n)$ is closed and compact. Let $h : GL_n(\mathbf{R}) \to O(n)$ be defined by Gram-Schmidt. Show $H : GL_n(\mathbf{R}) \times [0, 1] \to GL_n(\mathbf{R})$ defined by $H(A, t) = (1 - t)A + th(A)$ is a deformation retract of $GL_n(R)$ to $O(n)$.

Diagonalization of Symmetric Matrices

We continue with the case $F = \mathbf{R}$. Our goals are to prove that, if A is a symmetric matrix, all of its eigenvalues are real and that $\exists$ an orthogonal matrix C such that $C^{-1}AC$ is diagonal. As background, we first note that symmetric is the same as self-adjoint.

Theorem Suppose $A \in \mathbf{R}_n$ and $u, v \in \mathbf{R}^n$. Then $(A^t u) \cdot v = u \cdot (Av)$.

Proof If $y, z \in \mathbf{R}^n$, then the dot product $y \cdot z$, is the matrix product $y^t z$, and matrix multiplication is associative. Thus $(A^t u) \cdot v = (u^t A)v = u^t(Av) = u \cdot (Av)$.

Definition Suppose $A \in \mathbf{R}_n$. A is said to be *symmetric* provided $A^t = A$. Note that any diagonal matrix is symmetric. A is said to be *self-adjoint* if $(Au) \cdot v = u \cdot (Av)$ for all $u, v \in \mathbf{R}^n$. The next theorem is just an exercise using the previous theorem.

Theorem A is symmetric iff A is self-adjoint.

Theorem Suppose $A \in \mathbf{R}_n$ is symmetric. Then $\exists$ real numbers $\lambda_1, .., \lambda_n$ (not necessarily distinct) such that $CP_A(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$. That is, all the eigenvalues of A are real.

Proof We know $CP_A(x)$ factors into linears over $\mathbf{C}$. If $\mu = a + bi$ is a complex number, its conjugate is defined by $\bar{\mu} = a - bi$. If $h : \mathbf{C} \to \mathbf{C}$ is defined by $h(\mu) = \bar{\mu}$, then h is a ring isomorphism which is the identity on $\mathbf{R}$. If $w = (a_{i,j})$ is a complex matrix or vector, its conjugate is defined by $\bar{w} = (\bar{a}_{i,j})$. Since $A \in \mathbf{R}_n$ is a real symmetric matrix, $A = A^t = \bar{A}^t$. Now suppose λ is a complex eigenvalue of A and $v \in \mathbf{C}^n$ is an eigenvector with $Av = \lambda v$. Then $\lambda(v^t\bar{v}) = (\lambda v)^t\bar{v} = (Av)^t\bar{v} = (v^tA)\bar{v} = v^t(A\bar{v}) = v^t(\overline{Av}) = v^t(\overline{\lambda v}) = \bar{\lambda}(v^t\bar{v})$. Thus $\lambda = \bar{\lambda}$ and $\lambda \in \mathbf{R}$. Or you can define a complex inner product on $\mathbf{C}^n$ by $(w \cdot v) = w^t\bar{v}$. The proof then reads as $\lambda(v \cdot v) = (\lambda v \cdot v) = (Av \cdot v) = (v \cdot Av) = (v \cdot \lambda v) = \bar{\lambda}(v \cdot v)$. Either way, λ is a real number.

We know that eigenvectors belonging to distinct eigenvalues are linearly independent. For symmetric matrices, we show more, namely that they are perpendicular.

Theorem Suppose A is symmetric, $\lambda_1, \lambda_2 \in \mathbf{R}$ are distinct eigenvalues of A, and $Au = \lambda_1 u$ and $Av = \lambda_2 v$. Then $u \cdot v = 0$.

Proof $\lambda_1(u \cdot v) = (Au) \cdot v = u \cdot (Av) = \lambda_2(u \cdot v)$.

Review Suppose $A \in \mathbf{R}_n$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is defined by $f(B) = AB$. Then A represents f w.r.t. the canonical orthonormal basis. Let $S = \{v_1, .., v_n\}$ be another basis and $C \in \mathbf{R}_n$ be the matrix with v_i as column i. Then $C^{-1}AC$ is the matrix representing f w.r.t. S. Now S is an orthonormal basis iff C is an orthogonal matrix.

Summary Representing f w.r.t. an orthonormal basis is the same as conjugating A by an orthogonal matrix.

Theorem Suppose $A \in \mathbf{R}_n$ and $C \in O(n)$. Then A is symmetric iff $C^{-1}AC$ is symmetric.

Proof Suppose A is symmetric. Then $(C^{-1}AC)^t = C^tA(C^{-1})^t = C^{-1}AC$.

The next theorem has geometric and physical implications, but for us, just the incredibility of it all will suffice.

Theorem If $A \in \mathbf{R}_n$, the following are equivalent.

1) A is symmetric.
2) $\exists\ C \in O(n)$ such that $C^{-1}AC$ is diagonal.

Proof By the previous theorem, 2) $\Rightarrow$ 1). Show 1) $\Rightarrow$ 2). Suppose A is a symmetric 2×2 matrix. Let λ be an eigenvalue for A and $\{v_1, v_2\}$ be an orthonormal basis for $\mathbf{R}^2$ with $Av_1 = \lambda v_1$. Then w.r.t this basis, the transformation determined by A is represented by $\begin{pmatrix} \lambda & b \\ 0 & d \end{pmatrix}$. Since this matrix is symmetric, $b = 0$.

Now suppose by induction that the theorem is true for symmetric matrices in $\mathbf{R}_t$ for $t < n$, and suppose A is a symmetric $n \times n$ matrix. Denote by $\lambda_1, .., \lambda_k$ the distinct eigenvalues of A, $k \leq n$. If $k = n$, the proof is immediate, because then there is a basis of eigenvectors of length 1, and they must form an orthonormal basis. So suppose $k < n$. Let $v_1, .., v_k$ be eigenvectors for $\lambda_1, .., \lambda_k$ with each $\| v_i \| = 1$. They may be extended to an orthonormal basis $v_1, .., v_n$. With respect to this basis, the transformation determined by A is represented by $\begin{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix} & (B) \\ (0) & (D) \end{pmatrix}$.

Since this is a symmetric matrix, $B = 0$ and D is a symmetric matrix of smaller size. By induction, $\exists$ an orthogonal C such that $C^{-1}DC$ is diagonal. Thus conjugating by $\begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix}$ makes the entire matrix diagonal.

This theorem is so basic we state it again in different terminology. If V is an IPS, a linear transformation $f : V \to V$ is said to be self-adjoint provided $(u \cdot f(v)) = (f(u) \cdot v)$ for all $u, v \in V$.

Theorem If V is an n-dimensional IPS and $f : V \to V$ is a linear transformation, then the following are equivalent.

1) f is self-adjoint.
2) $\exists$ an orthonormal basis $\{v_1, ..., v_n\}$ for V with each v_i an eigenvector of f.

Exercise Let $A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$. Find an orthogonal C such that $C^{-1}AC$ is diagonal. Do the same for $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

Exercise Suppose $A, D \in \mathbf{R}_n$ are symmetric. Under what conditions are A and D similar? Show that, if A and D are similar, $\exists$ an orthogonal C such that $D = C^{-1}AC$.

Exercise Suppose V is an n-dimensional real vector space. We know that V is isomorphic to $\mathbf{R}^n$. Suppose f and g are isomorphisms from V to $\mathbf{R}^n$ and A is a subset of V. Show that $f(A)$ is an open subset of $\mathbf{R}^n$ iff $g(A)$ is an open subset of $\mathbf{R}^n$. This shows that V, an algebraic object, has a god-given topology. Of course, if V has an inner product, it automatically has a metric, and this metric will determine that same topology. Finally, suppose V and W are finite-dimensional real vector spaces and $h : V \to W$ is a linear transformation. Show that h is continuous.

Exercise Define $E : \mathbf{C}_n \to \mathbf{C}_n$ by $E(A) = e^A = I + A + (1/2!)A^2 + \cdots$. This series converges and thus E is a well defined function. If $AB = BA$, then $E(A + B) = E(A)E(B)$. Since A and $-A$ commute, $I = E(\underline{0}) = E(A - A) = E(A)E(-A)$, and thus $E(A)$ is invertible with $E(A)^{-1} = E(-A)$. Furthermore $E(A^t) = E(A)^t$, and if C is invertible, $E(C^{-1}AC) = C^{-1}E(A)C$. Now use the results of this section to prove the statements below. (For part 1, assume the Jordan form, i.e., assume any $A \in \mathbf{C}_n$ is similar to a lower triangular matrix.)

1) If $A \in \mathbf{C}_n$, then $\mid e^A \mid = e^{\text{trace}(A)}$. Thus if $A \in \mathbf{R}_n$, $\mid e^A \mid = 1$ iff $\text{trace}(A) = 0$.

2) $\exists$ a non-zero matrix $N \in \mathbf{R}_2$ with $e^N = I$.

3) If $N \in \mathbf{R}_n$ is symmetric, then $e^N = I$ iff $N = \underline{0}$.

4) If $A \in \mathbf{R}_n$ and $A^t = -A$, then $e^A \in O(n)$.