Chapter 4

Matrices and Matrix Rings

We first consider matrices in full generality, i.e., over an arbitrary ring R. However, after the first few pages, it will be assumed that R is commutative. The topics, such as invertible matrices, transpose, elementary matrices, systems of equations, and determinant, are all classical. The highlight of the chapter is the theorem that a square matrix is a unit in the matrix ring iff its determinant is a unit in the ring. This chapter concludes with the theorem that similar matrices have the same determinant, trace, and characteristic polynomial. This will be used in the next chapter to show that an endomorphism on a finitely generated vector space has a well defined determinant, trace, and characteristic polynomial.

Definition Suppose R is a ring and m and n are positive integers. Let $R_{m,n}$ be the collection of all $m \times n$ matrices

$$A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \quad \text{where each entry } \ a_{i,j} \in R.$$

A matrix may be viewed as m n-dimensional row vectors or as n m-dimensional column vectors. A matrix is said to be *square* if it has the same number of rows as columns. Square matrices are so important that they have a special notation, $R_n = R_{n,n}$. R^n is defined to be the additive abelian group $R \times R \times \cdots \times R$. To emphasize that R^n does not have a ring structure, we use the "sum" notation, $R^n = R \oplus R \oplus \cdots \oplus R$. Our convention is to write elements of R^n as column vectors, i.e., to identify R^n with $R_{n,1}$. If the elements of R^n are written as row vectors, R^n is identified with $R_{1,n}$.

Addition of matrices To "add" two matrices, they must have the same number of rows and the same number of columns, i.e., addition is a binary operation $R_{m,n} \times R_{m,n} \to R_{m,n}$. The addition is defined by $(a_{i,j}) + (b_{i,j}) = (a_{i,j} + b_{i,j})$, i.e., the i, j term of the sum is the sum of the i, j terms. The following theorem is just an observation.

Theorem $R_{m,n}$ is an additive abelian group. Its "zero" is the matrix $\underline{0} = \underline{0}_{m,n}$ all of whose terms are zero. Also $-(a_{i,j}) = (-a_{i,j})$. Furthermore, as additive groups, $R_{m,n} \approx R^{mn}$.

Scalar multiplication An element of R is called a *scalar*. A matrix may be "multiplied" on the right or left by a scalar. Right scalar multiplication is defined by $(a_{i,j})c = (a_{i,j} \cdot c)$. It is a function $R_{m,n} \times R \to R_{m,n}$. Note in particular that scalar multiplication is defined on R^n. Of course, if R is commutative, there is no distinction between right and left scalar multiplication.

Theorem Suppose $A, B \in R_{m,n}$ and $c, d \in R$. Then

$$\begin{aligned} (A+B)c &= Ac + Bc \\ A(c+d) &= Ac + Ad \\ A(cd) &= (Ac)d \end{aligned}$$

and

$$A\underline{1} = A$$

This theorem is entirely transparent. In the language of the next chapter, it merely states that $R_{m,n}$ is a right module over the ring R.

Multiplication of Matrices The matrix product AB is defined iff the number of columns of A is equal to the number of rows of B. The matrix AB will have the same number of rows as A and the same number of columns as B, i.e., multiplication is a function $R_{m,n} \times R_{n,p} \to R_{m,p}$. The product $(a_{i,j})(b_{i,j})$ is defined to be the matrix whose (s, t) term is $a_{s,1} \cdot b_{1,t} + \cdots + a_{s,n} \cdot b_{n,t}$, i.e., the dot product of row s of A with column t of B.

Exercise Consider real matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $U = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $W = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Find the matrices AU, UA, AV, VA, AW, and WA.

Definition The *identity matrix* $I_n \in R_n$ is the square matrix whose diagonal terms are $\underline{1}$ and whose off-diagonal terms are $\underline{0}$.

Theorem Suppose $A \in R_{m,n}$.
1) $\underline{0}_{p,m}A = \underline{0}_{p,n} \quad A\underline{0}_{n,p} = \underline{0}_{m,p}$
2) $I_m A = A = AI_n$

Theorem (The distributive laws) $(A+B)C = AC + BC$ and $C(A+B) = CA + CB$ whenever the operations are defined.

Theorem (The associative law for matrix multiplication) Suppose $A \in R_{m,n}$, $B \in R_{n,p}$, and $C \in R_{p,q}$. Then $(AB)C = A(BC)$. Note that $ABC \in R_{m,q}$.

Proof We must show that the (s,t) terms are equal. The proof involves writing it out and changing the order of summation. Let $(x_{i,j}) = AB$ and $(y_{i,j}) = BC$. Then the (s,t) term of $(AB)C$ is $\sum_i x_{s,i}c_{i,t} = \sum_i \Big(\sum_j a_{s,j}b_{j,i}\Big)c_{i,t} = \sum_{i,j} a_{s,j}b_{j,i}c_{i,t} = \sum_j a_{s,j}\Big(\sum_i b_{j,i}c_{i,t}\Big) = \sum_j a_{s,j}y_{j,t}$ which is the (s,t) term of $A(BC)$.

Theorem For each ring R and integer $n \geq 1$, R_n is a ring.

Proof This elegant little theorem is immediate from the theorems above. The units of R_n are called *invertible* or *non-singular* matrices. They form a group under multiplication called the *general linear group* and denoted by $GL_n(R) = (R_n)^*$.

Exercise Recall that if A is a ring and $a \in A$, then aA is right ideal of A. Let $A = R_2$ and $a = (a_{i,j})$ where $a_{1,1} = \underline{1}$ and the other entries are $\underline{0}$. Find aR_2 and R_2a. Show that the only ideal of R_2 containing a is R_2 itself.

Multiplication by blocks Suppose $A, E \in R_n$, $B, F \in R_{n,m}$, $C, G \in R_{m,n}$, and $D, H \in R_m$. Then multiplication in R_{n+m} is given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{pmatrix}.$$

Transpose

Notation *For the remainder of this chapter on matrices, suppose R is a commutative ring.* Of course, for $n > 1$, R_n is non-commutative.

Transpose is a function from $R_{m,n}$ to $R_{n,m}$. If $A \in R_{m,n}$, $A^t \in R_{n,m}$ is the matrix whose (i,j) term is the (j,i) term of A. So row i (column i) of A becomes column i (row i) of A^t. If A is an n-dimensional row vector, then A^t is an n-dimensional column vector. If A is a square matrix, A^t is also square.

Theorem

1) $(A^t)^t = A$
2) $(A+B)^t = A^t + B^t$
3) If $c \in R$, $(Ac)^t = A^t c$
4) $(AB)^t = B^t A^t$
5) If $A \in R_n$, then A is invertible iff A^t is invertible. In this case $(A^{-1})^t = (A^t)^{-1}$.

Proof of 5) Suppose A is invertible. Then $I = I^t = (AA^{-1})^t = (A^{-1})^t A^t$.

Exercise Characterize those invertible matrices $A \in \mathbf{R}_2$ which have $A^{-1} = A^t$. Show that they form a subgroup of $GL_2(\mathbf{R})$.

Triangular Matrices

If $A \in R_n$, then A is *upper (lower) triangular* provided $a_{i,j} = \underline{0}$ for all $i > j$ (all $j > i$). A is *strictly upper (lower) triangular* provided $a_{i,j} = \underline{0}$ for all $i \geq j$ (all $j \geq i$). A is *diagonal* if it is upper and lower triangular, i.e., $a_{i,j} = \underline{0}$ for all $i \neq j$. Note that if A is upper (lower) triangular, then A^t is lower (upper) triangular.

Theorem If $A \in R_n$ is strictly upper (or lower) triangular, then $A^n = \underline{0}$.

Proof The way to understand this is just multiply it out for $n = 2$ and $n = 3$. The geometry of this theorem will become transparent later in Chapter 5 when the matrix A defines an R-module endomorphism on R^n (see page 93).

Definition If T is any ring, an element $t \in T$ is said to be *nilpotent* provided $\exists n$ such that $t^n = 0$. In this case, $(\underline{1} - t)$ is a unit with inverse $\underline{1} + t + t^2 + \cdots + t^{n-1}$. Thus if $T = R_n$ and B is a nilpotent matrix, $I - B$ is invertible.

Exercise Let $R = \mathbf{Z}$. Find the inverse of $\begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$.

Exercise Suppose $A = \begin{pmatrix} a_1 & & & \\ & a_2 & & 0 \\ & & \cdot & \\ & 0 & & \cdot \\ & & & & a_n \end{pmatrix}$ is a diagonal matrix, $B \in R_{m,n}$, and $C \in R_{n,p}$. Show that BA is obtained from B by multiplying column i of B by a_i. Show AC is obtained from C by multiplying row i of C by a_i. Show A is a unit in R_n iff each a_i is a unit in R.

Scalar matrices A *scalar* matrix is a diagonal matrix for which all the diagonal terms are equal, i.e., a matrix of the form cI_n. The map $R \to R_n$ which sends c to cI_n is an injective ring homomorphism, and thus we may consider R to be a subring of R_n. Multiplying by a scalar is the same as multiplying by a scalar matrix, and thus scalar matrices commute with everything, i.e., if $B \in R_n$, $(cI_n)B = cB = Bc = B(cI_n)$. Recall we are assuming R is a commutative ring.

Exercise Suppose $A \in R_n$ and for each $B \in R_n$, $AB = BA$. Show A is a scalar matrix. For $n > 1$, this shows how non-commutative R_n is.

Elementary Operations and Elementary Matrices

Suppose R is a commutative ring and A is a matrix over R. There are 3 types of elementary row and column operations on the matrix A. A need not be square.

Type 1	Multiply row i by some unit $a \in R$.	Multiply column i by some unit $a \in R$.
Type 2	Interchange row i and row j.	Interchange column i and column j.
Type 3	Add a times row j to row i where $i \neq j$ and a is any element of R.	Add a times column i to column j where $i \neq j$ and a is any element of R.

Elementary Matrices Elementary matrices are square and invertible. There are three types. They are obtained by performing row or column operations on the identity matrix.

Type 1
$$B = \begin{pmatrix} 1 & & & & & \\ & 1 & & & 0 & \\ & & a & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{pmatrix}$$
where a is a unit in R.

Type 2
$$B = \begin{pmatrix} 1 & & & & & \\ & 0 & & & 1 & \\ & & 1 & & & \\ & & & 1 & & \\ & 1 & & & 0 & \\ & & & & & 1 \end{pmatrix}$$

Type 3
$$B = \begin{pmatrix} 1 & & & & & \\ & 1 & & & a_{i,j} & \\ & & 1 & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{pmatrix}$$
where $i \neq j$ and $a_{i,j}$ is any element of R.

In type 1, all the off-diagonal elements are zero. In type 2, there are two non-zero off-diagonal elements. In type 3, there is at most one non-zero off-diagonal element, and it may be above or below the diagonal.

Exercise Show that if B is an elementary matrix of type 1,2, or 3, then B is invertible and B^{-1} is an elementary matrix of the same type.

The following theorem is handy when working with matrices.

Theorem Suppose A is a matrix. It need not be square. To perform an elementary row (column) operation on A, perform the operation on an identity matrix to obtain an elementary matrix B, and multiply on the left (right). That is, $BA =$ row operation on A and $AB =$ column operation on A. (See the exercise on page 54.)

Exercise Suppose F is a field and $A \in F_{m,n}$.

1) Show $\exists$ invertible matrices $B \in F_m$ and $C \in F_n$ such that $BAC = (d_{i,j})$ where $d_{1,1} = \cdots = d_{t,t} = \underline{1}$ and all other entries are $\underline{0}$. The integer t is called the *rank* of A. (See page 89 of Chapter 5.)

2) Suppose $A \in F_n$ is invertible. Show A is the product of elementary matrices.

3) A matrix T is said to be in *row echelon* form if, for each $1 \leq i < m$, the first non-zero term of row $(i+1)$ is to the right of the first non-zero term of row i. Show $\exists$ an invertible matrix $B \in F_m$ such that BA is in row echelon form.

4) Let $A = \begin{pmatrix} 3 & 11 \\ 0 & 4 \end{pmatrix}$ and $D = \begin{pmatrix} 3 & 11 \\ 1 & 4 \end{pmatrix}$. Write A and D as products of elementary matrices over $\mathbf{Q}$. Is it possible to write them as products of elementary matrices over $\mathbf{Z}$?

For 1), perform row and column operations on A to reach the desired form. This shows the matrices B and C may be selected as products of elementary matrices. Part 2) also follows from this procedure. For part 3), use only row operations. Notice that if T is in row-echelon form, the number of non-zero rows is the rank of T.

Systems of Equations

Suppose $A = (a_{i,j}) \in R_{m,n}$ and $C = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_m \end{pmatrix} \in R^m = R_{m,1}$. The system

$$\begin{array}{rcl} a_{1,1}x_1 + \cdots + a_{1,n}x_n = & & c_1 \\ \vdots \qquad\qquad\quad \vdots & & \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n = & & c_m \end{array}$$

of m equations in n unknowns, can be written as one matrix equation in one unknown, namely as $(a_{i,j}) \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ c_m \end{pmatrix}$ or $AX = C$.

Define $f : R^n \to R^m$ by $f(D) = AD$. Then f is a group homomorphism and also $f(Dc) = f(D)c$ for any $c \in R$. In the language of the next chapter, this says that f is an R-module homomorphism. The next theorem summarizes what we already know about solutions of linear equations in this setting.

Theorem

1) $AX = \underline{0}$ is called the *homogeneous equation*. Its solution set is $\ker(f)$.

2) $AX = C$ has a solution iff $C \in \text{image}(f)$. If $D \in R^n$ is one solution, the solution set $f^{-1}(C)$ is the coset $D + \ker(f)$ in R^n. (See part 7 of the theorem on homomorphisms in Chapter 2, page 28.)

3) Suppose $B \in R_m$ is invertible. Then $AX = C$ and $(BA)X = BC$ have the same set of solutions. Thus we may perform any row operation on both sides of the equation and not change the solution set.

4) If $m = n$ and $A \in R_m$ is invertible, then $AX = C$ has the unique solution $X = A^{-1}C$.

The geometry of systems of equations over a field will not become really transparent until the development of linear algebra in Chapter 5.

Determinants

The concept of determinant is one of the most amazing in all of mathematics. The proper development of this concept requires a study of multilinear forms, which is given in Chapter 6. In this section we simply present the basic properties.

For each $n \geq 1$ and each commutative ring R, determinant is a function from R_n to R. For $n = 1$, $| (a) | = a$. For $n = 2$, $\left| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = ad - bc$.

Definition Let $A = (a_{i,j}) \in R_n$. If σ is a permutation on $\{1, 2, ..., n\}$, let $\text{sign}(\sigma) = 1$ if σ is an even permutation, and $\text{sign}(\sigma) = -1$ if σ is an odd permutation. The *determinant* is defined by $| A | = \sum_{\text{all } \sigma} \text{sign}(\sigma)\, a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$. Check that for $n = 2$, this agrees with the definition above. (Note that here we are writing the permutation functions as $\sigma(i)$ and not as $(i)\sigma$.)

For each σ, $a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$ contains exactly one factor from each row and one factor from each column. Since R is commutative, we may rearrange the factors so that the first comes from the first column, the second from the second column, etc. This means that there is a permutation τ on $\{1, 2, \ldots, n\}$ such that $a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} = a_{\tau(1),1} \cdots a_{\tau(n),n}$. We wish to show that $\tau = \sigma^{-1}$ and thus $\text{sign}(\sigma) = \text{sign}(\tau)$. To reduce the abstraction, suppose $\sigma(2) = 5$. Then the first expression will contain the factor $a_{2,5}$. In the second expression, it will appear as $a_{\tau(5),5}$, and so $\tau(5) = 2$. Anyway, τ is the inverse of σ and thus there are two ways to define determinant. It follows that the determinant of a matrix is equal to the determinant of its transpose.

Theorem $|A| = \sum_{\text{all } \sigma} \text{sign}(\sigma) a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} = \sum_{\text{all } \tau} \text{sign}(\tau) a_{\tau(1),1} \cdot a_{\tau(2),2} \cdots a_{\tau(n),n}$.

Corollary $|A| = |A^t|$.

You may view an $n \times n$ matrix A as a sequence of n column vectors or as a sequence of n row vectors. Here we will use column vectors. This means we write the matrix A as $A = (A_1, A_2, \ldots, A_n)$ where each $A_i \in R_{n,1} = R^n$.

Theorem If two columns of A are equal, then $|A| = \underline{0}$.

Proof For simplicity, assume the first two columns are equal, i.e., $A_1 = A_2$. Now $|A| = \sum_{\text{all } \tau} \text{sign}(\tau) a_{\tau(1),1} \cdot a_{\tau(2),2} \cdots a_{\tau(n),n}$ and this summation has $n!$ terms and $n!$ is an even number. Let γ be the transposition which interchanges one and two. Then for any τ, $a_{\tau(1),1} \cdot a_{\tau(2),2} \cdots a_{\tau(n),n} = a_{\tau\gamma(1),1} \cdot a_{\tau\gamma(2),2} \cdots a_{\tau\gamma(n),n}$. This pairs up the $n!$ terms of the summation, and since $\text{sign}(\tau) = -\text{sign}(\tau\gamma)$, these pairs cancel in the summation. Therefore $|A| = \underline{0}$.

Theorem Suppose $1 \leq r \leq n$, $C_r \in R_{n,1}$, and $a, c \in R$. Then $|(A_1, \ldots, A_{r-1}, aA_r + cC_r, A_{r+1}, \ldots, A_n)| = a|(A_1, \ldots, A_n)| + c|(A_1, \ldots, A_{r-1}, C_r, A_{r+1}, \ldots, A_n)|$

Proof This is immediate from the definition of determinant and the distributive law of multiplication in the ring R.

Summary Determinant is a function $d : R_n \to R$. In the language used in the Appendix, the two previous theorems say that d is an alternating multilinear form. The next two theorems show that alternating implies skew-symmetric (see page 129).

Theorem Interchanging two columns of A multiplies the determinant by minus one.

Proof For simplicity, show that $|(A_2, A_1, A_3, \ldots, A_n)| = -|A|$. We know $\underline{0} = |(A_1 + A_2, A_1 + A_2, A_3, \ldots, A_n)| = |(A_1, A_1, A_3, \ldots, A_n)| + |(A_1, A_2, A_3, \ldots, A_n)| + |(A_2, A_1, A_3, \ldots, A_n)| + |(A_2, A_2, A_3, \ldots, A_n)|$. Since the first and last of these four terms are zero, the result follows.

Theorem If τ is a permutation of $(1, 2, \ldots, n)$, then $|A| = \text{sign}(\tau)|(A_{\tau(1)}, A_{\tau(2)}, \ldots, A_{\tau(n)})|$.

Proof The permutation τ is the finite product of transpositions.

Exercise Rewrite the four preceding theorems using rows instead of columns.

The following theorem is just a summary of some of the work done so far.

Theorem Multiplying any row or column of matrix by a scalar $c \in R$, multiplies the determinant by c. Interchanging two rows or two columns multiplies the determinant by -1. Adding c times one row to another row, or adding c times one column to another column, does not change the determinant. If a matrix has two rows equal or two columns equal, its determinant is zero. More generally, if one row is c times another row, or one column is c times another column, then the determinant is zero.

There are $2n$ ways to compute $|\,A\,|$; expansion by any row or expansion by any column. Let $M_{i,j}$ be the determinant of the $(n-1) \times (n-1)$ matrix obtained by removing row i and column j from A. Let $C_{i,j} = (-1)^{i+j} M_{i,j}$. $M_{i,j}$ and $C_{i,j}$ are called the (i, j) *minor* and *cofactor* of A. The following theorem is useful but the proof is a little tedious and should not be done as an exercise.

Theorem For any $1 \leq i \leq n$, $|\,A\,| = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n}$. For any $1 \leq j \leq n$, $|\,A| = a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}$. Thus if any row or any column is zero, the determinant is zero.

Exercise Let $A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$. The determinant of A is the sum of six terms.

Write out the determinant of A expanding by the first column and also expanding by the second row.

Theorem If A is an upper or lower triangular matrix, $|\,A\,|$ is the product of the diagonal elements. If A is an elementary matrix of type 2, $|\,A\,| = -1$. If A is an elementary matrix of type 3, $|\,A| = 1$.

Proof We will prove the first statement for upper triangular matrices. If $A \in R_2$ is an upper triangular matrix, then its determinant is the product of the diagonal elements. Suppose $n > 2$ and the theorem is true for matrices in R_{n-1}. Suppose $A \in R_n$ is upper triangular. The result follows by expanding by the first column.

An elementary matrix of type 3 is a special type of upper or lower triangular matrix, so its determinant is 1. An elementary matrix of type 2 is obtained from the identity matrix by interchanging two rows or columns, and thus has determinant -1.

Theorem (Determinant by blocks) Suppose $A \in R_n$, $B \in R_{n,m}$, and $D \in R_m$. Then the determinant of $\begin{pmatrix} A & B \\ O & D \end{pmatrix}$ is $|\,A||\,D\,|$.

Proof Expand by the first column and use induction on n.

The following remarkable theorem takes some work to prove. We assume it here without proof. (For the proof, see page 130 of the Appendix.)

Theorem The determinant of the product is the product of the determinants, i.e., if $A, B \in R_n$, $|\,AB\,| = |\,A\,||\,B\,|$. Thus $|\,AB\,| = |\,BA\,|$ and if C is invertible, $|\,C^{-1}AC\,| = |ACC^{-1}| = |\,A|$.

Corollary If A is a unit in R_n, then $|\,A|$ is a unit in R and $|\,A^{-1}| = |\,A|^{-1}$.

Proof $\underline{1} = |\,I\,| = |\,AA^{-1}| = |\,A||\,A^{-1}|\,.$

One of the major goals of this chapter is to prove the converse of the preceding corollary.

Classical adjoint Suppose R is a commutative ring and $A \in R_n$. The *classical adjoint* of A is $(C_{i,j})^t$, i.e., the matrix whose (j,i) term is the (i,j) cofactor. Before

we consider the general case, let's examine 2×2 matrices.

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $(C_{i,j}) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$ and so $(C_{i,j})^t = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Then $A(C_{i,j})^t = (C_{i,j})^t A = \begin{pmatrix} |A| & 0 \\ 0 & |A| \end{pmatrix} = |A|\ I$. Thus if $|A|$ is a unit in R, A is invertible and $A^{-1} = |A|^{-1}\ (C_{i,j})^t$. In particular, if $|A| = \underline{1}$, $A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Here is the general case.

Theorem If R is commutative and $A \in R_n$, then $A(C_{i,j})^t = (C_{i,j})^t A = |A|\ I$.

Proof We must show that the diagonal elements of the product $A(C_{i,j})^t$ are all $|A|$ and the other elements are $\underline{0}$. The (s,s) term is the dot product of row s of A with row s of $(C_{i,j})$ and is thus $|A|$ (computed by expansion by row s). For $s \neq t$, the (s,t) term is the dot product of row s of A with row t of $(C_{i,j})$. Since this is the determinant of a matrix with row s $=$ row t, the (s,t) term is $\underline{0}$. The proof that $(C_{i,j})^t A = |A| I$ is similar and is left as an exercise.

We are now ready for one of the most beautiful and useful theorems in all of mathematics.

Theorem Suppose R is a commutative ring and $A \in R_n$. Then A is a unit in R_n iff $|A|$ is a unit in R. (Thus if R is a field, A is invertible iff $|A| \neq \underline{0}$.) If A is invertible, then $A^{-1} = |A|^{-1}\ (C_{i,j})^t$. Thus if $|A| = \underline{1}$, $A^{-1} = (C_{i,j})^t$, the classical adjoint of A.

Proof This follows immediately from the preceding theorem.

Exercise Show that any right inverse of A is also a left inverse. That is, suppose $A, B \in R_n$ and $AB = I$. Show A is invertible with $A^{-1} = B$, and thus $BA = I$.

Similarity

Suppose $A, B \in R_n$. B is said to be *similar* to A if $\exists$ an invertible $C \in R_n$ such that $B = C^{-1}AC$, i.e., B is similar to A iff B is a *conjugate* of A.

Theorem B is similar to B.

B is similar to A iff A is similar to B.

If D is similar to B and B is similar to A, then D is similar to A.

"Similarity" is an equivalence relation on R_n.

Proof This is a good exercise using the definition.

Theorem Suppose A and B are similar. Then $|A| = |B|$ and thus A is invertible iff B is invertible.

Proof Suppose $B = C^{-1}AC$. Then $|B| = |C^{-1}AC| = |ACC^{-1}| = |A|$.

Trace Suppose $A = (a_{i,j}) \in R_n$. Then the *trace* is defined by $\text{trace}(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}$. That is, the trace of A is the sum of its diagonal terms.

One of the most useful properties of trace is $\text{trace}(AB) = \text{trace}(BA)$ whenever AB and BA are defined. For example, suppose $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_n)^t$. Then AB is the scalar $a_1b_1 + \cdots + a_nb_n$ while BA is the $n \times n$ matrix (b_ia_j). Note that $\text{trace}(AB) = \text{trace}(BA)$. Here is the theorem in full generality.

Theorem Suppose $A \in R_{m,n}$ and $B \in R_{n,m}$. Then AB and BA are square matrices with $\text{trace}(AB) = \text{trace}(BA)$.

Proof This proof involves a change in the order of summation. By definition, $\text{trace}(AB) = \sum_{1 \le i \le m} a_{i,1}b_{1,i} + \cdots + a_{i,n}b_{n,i} = \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} a_{i,j}b_{j,i} = \sum_{1 \le j \le n} b_{j,1}a_{1,j} + \cdots + b_{j,m}a_{m,j} =$ $\text{trace}(BA)$.

Theorem If $A, B \in R_n$, $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$ and $\text{trace}(AB) = \text{trace}(BA)$.

Proof The first part of the theorem is immediate, and the second part is a special case of the previous theorem.

Theorem If A and B are similar, then $\text{trace}(A) = \text{trace}(B)$.

Proof $\text{trace}(B) = \text{trace}(C^{-1}AC) = \text{trace}(ACC^{-1}) = \text{trace}(A)$.

Summary Determinant and trace are functions from R_n to R. Determinant is a multiplicative homomorphism and trace is an additive homomorphism. Furthermore $|\,AB\,| = |\,BA\,|$ and $\text{trace}(AB) = \text{trace}(BA)$. If A and B are similar, $|\,A\,| = |\,B\,|$ and $\text{trace}(A) = \text{trace}(B)$.

Exercise Suppose $A \in R_n$ and $a \in R$. Find $|aA|$ and $\text{trace}(aA)$.

Characteristic polynomials If $A \in R_n$, the *characteristic polynomial* $CP_A(x) \in R[x]$ is defined by $CP_A(x) = |\,(xI - A)\,|$. Any $\lambda \in R$ which is a root of $CP_A(x)$ is called a *characteristic root* of A.

Theorem $CP_A(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$ where $\text{trace}(A) = -a_{n-1}$ and $|\,A\,| = (-1)^n a_0$.

Proof This follows from a direct computation of the determinant.

Theorem If A and B are similar, then they have the same characteristic polynomials.

Proof Suppose $B = C^{-1}AC$. $CP_B(x) = |\,(xI - C^{-1}AC)\,| = |\,C^{-1}(xI - A)C\,| = |\,(xI - A)\,| = CP_A(x)$.

Exercise Suppose R is a commutative ring, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a matrix in R_2, and $CP_A(x) = a_0 + a_1x + x^2$. Find a_0 and a_1 and show that $a_0I + a_1A + A^2 = \underline{0}$, i.e., show A satisfies its characteristic polynomial. In other words, $CP_A(A) = \underline{0}$.

Exercise Suppose F is a field and $A \in F_2$. Show the following are equivalent.

1) $A^2 = \underline{0}$.
2) $|\,A\,| = \text{trace}(A) = \underline{0}$.
3) $CP_A(x) = x^2$.
4) $\exists$ an elementary matrix C such that $C^{-1}AC$ is strictly upper triangular.

Note This exercise is a special case of a more general theorem. A square matrix over a field is nilpotent iff all its characteristic roots are $\underline{0}$ iff it is similar to a strictly upper triangular matrix. This remarkable result cannot be proved by matrix theory alone, but depends on linear algebra (see pages 93, 94, and 98).