CHAPTER 0

Review and miscellanea

0.0 Introduction

†In this initial chapter we summarize many useful concepts and facts, some of which provide a foundation for the material in the rest of the book. Much of this material is included in a typical elementary course in linear algebra, but we also include some useful items that are not commonly found elsewhere and are not used explicitly in our exposition. Thus, the reader may use this chapter for a short review prior to beginning the main part of the book; later it can serve as a convenient reference. We use it to set basic notation and give definitions that are used without further comment in later chapters. We assume that the reader is already familiar with the elementary concepts of linear algebra and with mechanical aspects of matrix manipulations, such as matrix multiplication and addition.

0.1 Vector spaces

A vector space is the fundamental setting for matrix analysis.

0.1.1 Scalar field. Underlying a vector space is its *field*, or set of scalars. For our purposes, that underlying field is almost always the real numbers $\mathbf{R}$ or the complex numbers $\mathbf{C}$ (see Appendix **A**) under the usual addition and multiplication, but it could be the rational numbers, the integers modulo a specified prime number, or some other field. When the field is unspecified, we denote it by the symbol $\mathbf{F}$. To qualify as a field, a set must be closed under two binary operations: "addition" and "multiplication"; both operations must be associative and commutative and each must have an identity element in the set; inverses must exist in the set for all elements under addition and for all elements except the additive identity under multiplication; multiplication must be distributive over addition.

0.1.2 Vector spaces. A *vector space* V over a field $\mathbf{F}$ is a set V of objects (called *vectors*) that is closed under a binary operation ("addition") that is associative and commutative and has an identity (the *zero vector*, denoted by 0) and additive inverses in the set. The set is also closed under an operation of "scalar multiplication" of the vectors by elements of the scalar field $\mathbf{F}$, with the following properties for all $a, b \in \mathbf{F}$ and all $x, y \in V$: $a(x+y) = ax + ay$, $(a+b)x = ax + bx$, $a(bx) = (ab)x$, and $ex = x$ for the multiplicative identity $e \in \mathbf{F}$.

For a given field $\mathbf{F}$ and a given positive integer n the set $\mathbf{F}^n$ of n-tuples with entries from $\mathbf{F}$ forms a vector space over $\mathbf{F}$ under entry-wise addition in $\mathbf{F}^n$. Our convention is that *elements of* $\mathbf{F}^n$ *are always presented as column vectors*; we often call them n*-vectors*. The special cases $\mathbf{R}^n$ and $\mathbf{C}^n$ are the basic vector spaces of this book. The set of polynomials with real or with complex coefficients (of no more than a specified degree *or* of arbitrary degree) and the set of real or complex valued continuous functions or arbitrary functions on subsets of $\mathbf{R}$ or $\mathbf{C}$ are also examples of vector spaces (over $\mathbf{R}$ or $\mathbf{C}$).

0.1.3 Subspaces, span, and linear combinations. A *subspace* of a vector space V over a field $\mathbf{F}$ is a subset of V that is, by itself, a vector space over $\mathbf{F}$ using the same operations of vector addition and scalar multiplication as in V. A subset of V is a subspace precisely when it is closed under these two operations. For example, $\{[a, b, 0]^T : a, b \in \mathbf{R}\}$ is a subspace of $\mathbf{R}^3$ [see (0.2.5) for the transpose notation]. An intersection of subspaces is always a subspace; a union of subspaces need not be a subspace. The subsets $\{0\}$ and V are always subspaces of V, so they are often called *trivial subspaces*; a subspace of V is said to be *nontrivial* if it is different from both $\{0\}$ and V. A subspace of V is said to be a *proper subspace* if it is not equal to V. We call $\{0\}$ the *zero vector space*. Since a vector space always contains the zero vector, a subspace cannot be empty.

If S is a subset of a vector space V over a field $\mathbf{F}$, $\operatorname{span} S$ is the intersection of all subspaces of V that contain S. If S is nonempty, then $\operatorname{span} S = \{a_1 v_1 + \cdots + a_k v_k : a_1, \ldots, a_k \in \mathbf{F}, v_1, \ldots, v_k \in S, \text{ and } k = 1, 2, \ldots\}$; if S is empty, it follows from the definition that $\operatorname{span} S = \{0\}$. Notice that $\operatorname{span} S$ is always a subspace even if S is not a subspace; S is said to *span* V if $\operatorname{span} S = V$.

A *linear combination* of vectors in a subset S of a vector space V over a field $\mathbf{F}$ is any expression of the form $a_1v_1 + \cdots + a_kv_k$ for some positive integer k, some scalars $a_1, \ldots, a_k \in \mathbf{F}$, and some vectors $v_1, \ldots, v_k \in S$. Thus, the span of a nonempty subset S of V is the set of all linear combinations of finitely many vectors in S.

Let S_1 and S_2 be subspaces of a vector space over a field $\mathbf{F}$. The *sum* of S_1 and S_2 is the subspace

$$\begin{aligned} S_1 + S_2 &= \operatorname{span}\{S_1 \cup S_2\} \\ &= \{\alpha x + \beta y : x \in S_1, y \in S_2, \text{ and } \alpha, \beta \in \mathbf{F}\} \end{aligned}$$

If $S_1 \cap S_2 = \{0\}$, we say that the sum of S_1 and S_2 is a *direct sum* and write it as $S_1 \oplus S_2$; every $z \in S_1 \oplus S_2$ can be written as $z = \alpha x + \beta y$ with $x \in S_1$, $y \in S_2$, and $\alpha, \beta \in \mathbf{F}$ in one and only one way.

0.1.4 Linear dependence and independence. Let S be a given set of vectors in a vector space V over a field $\mathbf{F}$. We say that the set S is *dependent* (more formally: *linearly dependent*) if for some integer $k \geq 1$ there are vectors $x_1, \ldots, x_k \in S$ and scalars $a_1, \ldots, a_k \in \mathbf{F}$ that are not all zero, such that $a_1x_1 + \cdots + a_kx_k = 0$. A set of two or more vectors is dependent if one of the vectors is a linear combination of some of the others. A set of two vectors is dependent if and only if one of the vectors is a scalar multiple of the other. For example, $\{[1,2,3]^T, [1,0,-1]^T, [2,2,2]^T\}$ is a dependent set in $\mathbf{R}^3$; the first vector is the third vector minus the second vector.

A subset of V is said to be *independent* (more formally: *linearly independent*) if it is not dependent. For example, $\{[1,2,3]^T, [1,0,-1]^T\}$ is an independent set in $\mathbf{R}^3$; neither vector is a scalar multiple of the other.

It is important to note that both independence and dependence are properties of *sets* of vectors. Any subset of an independent set is independent; any set that contains a dependent set is dependent. Since $\{0\}$ is a dependent set, any set that contains the zero vector is dependent. A set of vectors can be dependent, while any proper subset of it is independent. An empty subset of a vector space is not dependent, so it is independent.

0.1.5 Basis. An independent set that spans a vector space V is called a *basis* for V. Each element of V can be represented as a linear combination of vectors in a basis in one and only one way; this is no longer true if any element whatsoever is appended to or deleted from the basis. An independent set in V is a basis of V if and only if no set that properly contains it is independent.

A set that spans V is a basis for V if and only if no proper subset of it spans V. Every vector space has a basis. The empty set is a basis for the zero vector space.

0.1.6 Extension to a basis. Any independent set in a vector space V that is not already a basis may be extended to a basis of V; that is, if an independent set $S_1 \subset V$ is not a basis, then there is another independent set $S_2 \subset V$ such that $S_1 \cup S_2$ is a basis of V. A given independent set that is not a basis may always be extended to a basis in many different ways; for example, any 3-vector with nonzero third entry may be appended to the independent set $\{[1,0,0]^T, [0,1,0]^T\}$ to produce a basis of $\mathbf{R}^3$. The real vector space $C[0,1]$ of real-valued continuous functions on $[0,1]$ shows that a basis need not be finite; the infinite set of monomials $\{1, t, t^2, t^3, \ldots\}$ is an independent set in $C[0,1]$.

0.1.7 Dimension. If some basis of the vector space V consists of a finite number (a nonnegative integer) of elements, then all bases of V have the same number of elements; this common number is called the *dimension* of the vector space V, and is denoted by $\dim V$. In this event, V is said to be *finite-dimensional*; otherwise V is said to be *infinite-dimensional*. In the infinite-dimensional case (e.g., $C[0,1]$), there is a one-to-one correspondence between the elements of any two bases. The real vector space $\mathbf{R}^n$ has dimension n. The vector space $\mathbf{C}^n$ has dimension n over the field $\mathbf{C}$ but dimension $2n$ over the field $\mathbf{R}$. The basis $\{e_1, e_2, \ldots, e_n\}$ of $\mathbf{F}^n$ in which each n-vector e_i has a 1 as its ith entry and 0's elsewhere is called the *standard basis*.

It is convenient to say "V is an n-dimensional vector space" as a shorthand for "V is a finite-dimensional vector space whose dimension is n." Any subspace of an n-dimensional vector space is finite-dimensional; its dimension is strictly less than n if it is a proper subspace.

Let V be a finite-dimensional vector space and let S_1 and S_2 be two given subspaces of V. Then

$$\dim(S_1 \cap S_2) + \dim(S_1 + S_2) = \dim S_1 + \dim S_2 \qquad (0.1.7.1)$$

Rewriting this identity as

$$\begin{aligned} \dim(S_1 \cap S_2) &= \dim S_1 + \dim S_2 - \dim(S_1 + S_2) \qquad (0.1.7.2) \\ &\geq \dim S_1 + \dim S_2 - \dim V \end{aligned}$$

reveals the useful fact that if $\dim S_1 + \dim S_2 > \dim V$, then $S_1 \cap S_2$ contains a nonzero vector.

0.1.8 Isomorphism. If U and V are vector spaces over the same scalar field $\mathbf{F}$, and if $f : U \to V$ is an invertible function such that $f(ax+by) = af(x) + bf(y)$ for all $x, y \in U$ and all $a, b \in \mathbf{F}$, then f is said to be an *isomorphism* and U and V are said to be isomorphic ("same-structure"). Two finite-dimensional vector spaces over the same field are isomorphic if and only if they have the same dimension; thus, any n-dimensional vector space over $\mathbf{F}$ is isomorphic to $\mathbf{F}^n$. Any n-dimensional real vector space is, therefore, isomorphic to $\mathbf{R}^n$, and any n-dimensional complex vector space is isomorphic to $\mathbf{C}^n$. Specifically, if V is an n-dimensional vector space over a field $\mathbf{F}$ with specified basis $\mathcal{B} = \{x_1, \ldots, x_n\}$, then, since any element $x \in V$ may be written uniquely as $x = a_1x_1 + \cdots + a_nx_n$ in which each $a_i \in \mathbf{F}$, we may identify x with the n-vector $[x]_{\mathcal{B}} = [a_1, \ldots, a_n]^T$. For any basis $\mathcal{B}$, the mapping $x \to [x]_{\mathcal{B}}$ is an isomorphism between V and $\mathbf{F}^n$.

0.2 Matrices

The fundamental object of study here may be thought of in two important ways: as a rectangular array of scalars and as a linear transformation between two vector spaces, given specified bases for each space.

0.2.1 Rectangular arrays. A *matrix* is an m-by-n array of scalars from a field $\mathbf{F}$. If $m = n$, the matrix is said to be *square*. The set of all m-by-n matrices over $\mathbf{F}$ is denoted by $M_{m,n}(\mathbf{F})$, and $M_{n,n}(\mathbf{F})$ is often denoted by $M_n(\mathbf{F})$. The vector spaces $M_{n,1}(\mathbf{F})$ and $\mathbf{F}^n$ are identical. If $\mathbf{F} = \mathbf{C}$, then $M_n(\mathbf{C})$ is further abbreviated to M_n, and $M_{m,n}(\mathbf{C})$ to $M_{m,n}$. Matrices are typically denoted by capital letters and their scalar entries are typically denoted by doubly subscripted lower-case letters. For example, if

$$A = \begin{bmatrix} 2 & -\frac{3}{2} & 0 \\ -1 & \pi & 4 \end{bmatrix} = [a_{ij}]$$

then $A \in M_{2,3}(\mathbf{R})$ has entries $a_{11} = 2, a_{12} = -3/2, a_{13} = 0, a_{21} = -1, a_{22} = \pi, a_{23} = 4$. A *submatrix* of a given matrix is a rectangular array lying in specified subsets of the rows and columns of a given matrix. For example $[\pi\ 4]$ is a submatrix (lying in row 2 and columns 2 and 3) of A.

Suppose $A = [a_{ij}] \in M_{n,m}(\mathbf{F})$. The *main diagonal* of A is the list of entries $a_{11}, a_{22}, \ldots, a_{kk}$, in which $k = \min\{n, m\}$. It is sometimes convenient to express the main diagonal of A as a vector $\operatorname{diag}(A) = [a_{ii}]_{i=1}^k \in \mathbf{F}^k$. The p^{th} *superdiagonal* of A is the list $a_{1,p+1}, a_{2,p+2}, \ldots, a_{k,p+k}$, in which $k =$

$\min\{n, m-p\}$, $p = 0, 1, 2, \ldots, m-1$; the p^{th} *subdiagonal* of A is the list $a_{p+1,1}, a_{p+2,2}, \ldots, a_{p+\ell,\ell}$, in which $\ell = \min\{n-p, m\}$, $p = 0, 1, 2, \ldots, n-1$.

0.2.2 Linear transformations. Let U be an n-dimensional vector space and let V be an m-dimensional vector space, both over the same field $\mathbf{F}$; let $\mathcal{B}_U$ be a basis of U and let $\mathcal{B}_V$ be a basis of V. We may use the isomorphisms $x \to [x]_{\mathcal{B}_U}$ and $y \to [y]_{\mathcal{B}_V}$ to represent vectors in U and V as n-vectors and m-vectors over $\mathbf{F}$, respectively. A *linear transformation* is a function $T : U \to V$ such that $T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2)$ for any scalars $a_{1,}a_2$ and vectors x_1, x_2. A matrix $A \in M_{m,n}(\mathbf{F})$ corresponds to a linear transformation $T : U \to V$ in the following way: $y = T(x)$ if and only if $[y]_{\mathcal{B}_V} = A[x]_{\mathcal{B}_U}$. The matrix A is said to *represent the linear transformation* T (relative to the bases $\mathcal{B}_U$ and $\mathcal{B}_V$); the representing matrix A depends upon the bases chosen. When we study a matrix A, we realize that we are studying a linear transformation relative to a particular choice of bases, but explicit appeal to the bases is usually not necessary.

0.2.3 Vector spaces associated with a matrix or linear transformation. Any n-dimensional vector space over $\mathbf{F}$ may be identified with $\mathbf{F}^n$; we may think of $A \in M_{m,n}(\mathbf{F})$ as a linear transformation $x \to Ax$ from $\mathbf{F}^n$ to $\mathbf{F}^m$ (and also as an array). The *domain* of this linear transformation is $\mathbf{F}^n$; its *range* is $\operatorname{range} A = \{y \in \mathbf{F}^m : y = Ax\}$ for some $x \in \mathbf{F}^n$; its *null space* is $\operatorname{nullspace} A = \{x \in \mathbf{F}^n : Ax = 0\}$. The range of A is a subspace of $\mathbf{F}^m$, and the null space of A is a subspace of $\mathbf{F}^n$. The dimension of $\operatorname{nullspace} A$ is denoted by $\operatorname{nullity} A$; the dimension of $\operatorname{range} A$ is denoted by $\operatorname{rank} A$. These numbers are related by the *Rank-Nullity Theorem*

$$\begin{aligned} \dim(\operatorname{range} A) + \dim(\operatorname{nullspace} A) &= \\ \operatorname{rank} A + \operatorname{nullity} A &= n \end{aligned} \tag{0.2.3.1}$$

for $A \in M_{m,n}(\mathbf{F})$. The null space of A is a set of vectors in $\mathbf{F}^n$ whose entries satisfy m homogeneous linear equations.

0.2.4 Matrix operations. Matrix addition is defined entry-wise for arrays of the same dimensions and is denoted by $+$ ("$A + B$"). It corresponds to addition of linear transformations (relative to the same basis), and it inherits commutativity and associativity from the scalar field. The *zero matrix* (all entries are zero) is the additive identity, and $M_{m,n}(\mathbf{F})$ is a vector space over $\mathbf{F}$.

Matrix multiplication is denoted by juxtaposition ("AB") and corresponds to the composition of linear transformations. Therefore, it is defined only when $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{n,q}(\mathbf{F})$. It is associative, but not always commutative. For example,

$$\begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 3 & 8 \end{bmatrix}$$

The *identity matrix*

$$I = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \in M_n(\mathbf{F})$$

is the multiplicative identity in $M_n(\mathbf{F})$; its main diagonal entries are 1 and all other entries are 0. The identity matrix and any scalar multiple of it (a *scalar matrix*) commute with every matrix in $M_n(\mathbf{F})$; they are the only matrices that do so. Matrix multiplication is distributive over matrix addition.

The symbol 0 is used throughout the book to denote each of the following: the zero scalar of a field, the zero vector of a vector space, the zero n-vector in $\mathbf{F}^n$ (all entries equal to the zero scalar in $\mathbf{F}$), and the zero matrix in $M_{m,n}(\mathbf{F})$ (all entries equal to the zero scalar). The symbol I denotes the identity matrix of any size. If there is potential for confusion, we indicate the dimension of a zero or identity matrix with subscripts, e.g., $0_{p,q}$, 0_k, or I_k.

0.2.5 The transpose, conjugate transpose, and trace. If $A = [a_{ij}] \in M_{m,n}(\mathbf{F})$, the *transpose* of A, denoted by A^T, is the matrix in $M_{n,m}(\mathbf{F})$ whose i, j entry is a_{ji}; that is, rows are exchanged for columns and vice versa. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Of course, $(A^T)^T = A$. The *conjugate transpose* (sometime called the *adjoint* or *Hermitian adjoint*) of $A \in M_{m,n}(\mathbf{C})$, is denoted by A^* and defined by $A^* = \bar{A}^T$, in which $\bar{A}$ is the entry-wise conjugate. For example,

$$\begin{bmatrix} 1+i & 2-i \\ -3 & -2i \end{bmatrix}^* = \begin{bmatrix} 1-i & -3 \\ 2+i & 2i \end{bmatrix}$$

Both the transpose and the conjugate transpose obey the *reverse-order law:* $(AB)^* = B^*A^*$ and $(AB)^T = B^TA^T$. For the complex conjugate of a product, there is no reversing: $\overline{AB} = \bar{A}\bar{B}$. If x, y are real or complex vectors of the same size, then y^*x is a scalar and its conjugate transpose and complex conjugate are the same: $(y^*x)^* = \overline{y^*x} = x^*y = y^T\bar{x}$.

Many important classes of matrices are defined by identities involving the transpose or conjugate transpose. For example, $A \in M_n(\mathbf{F})$ is said to be *symmetric* if $A^T = A$, *skew symmetric* if $A^T = -A$, and *orthogonal* if $A^TA = I$; $A \in M_n(\mathbf{C})$ is said to be *Hermitian* if $A^* = A$, *skew Hermitian* if $A^* = -A$, *essentially Hermitian* if $e^{i\theta}A$ is Hermitian for some $\theta \in \mathbf{R}$, *unitary* if $A^*A = I$, and *normal* if $A^*A = AA^*$.

Each $A \in M_n(\mathbf{F})$ can be written in exactly one way as $A = S(A) + C(A)$, in which $S(A)$ is symmetric and $C(A)$ is skew-symmetric: $S(A) = \frac{1}{2}(A + A^T)$ is the *symmetric part* of A; $C(A) = \frac{1}{2}(A - A^T)$ is the *skew-symmetric part* of A.

Each $A \in M_{m,n}(\mathbf{C})$ can be written in exactly one way as $A = B + iC$, in which $B, C \in M_{m,n}(\mathbf{R})$: $B = \frac{1}{2}(A + \bar{A})$ is the *real part* of A; $C = \frac{1}{2i}(A - \bar{A})$ is the *imaginary part* of A.

Each $A \in M_n(\mathbf{C})$ can be written in exactly one way as $A = H(A) + iK(A)$, in which $H(A)$ and $K(A)$ are Hermitian: $H(A) = \frac{1}{2}(A + A^*)$ is the *Hermitian part* of A; $iK(A) = \frac{1}{2}(A - A^*)$ is the *skew-Hermitian part* of A.

The *trace* of a square matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is the sum of its main diagonal entries: $\operatorname{tr} A = a_{11} + \cdots + a_{nn}$. For any $A = [a_{ij}] \in M_{m,n}(\mathbf{C})$, $\operatorname{tr} AA^* = \sum_{i,j=1}^{m,n} |a_{ij}|^2$, so

$$\operatorname{tr} AA^* = 0 \text{ if and only if } A = 0 \tag{0.2.5.1}$$

0.2.6 Metamechanics of matrix multiplication. In addition to the conventional definition of matrix-vector and matrix-matrix multiplication, several alternative viewpoints can be useful.

a) If $A \in M_{m,n}(\mathbf{F})$, $x \in \mathbf{F}^n$, and $y \in \mathbf{F}^m$, then the (column) vector Ax is a linear combination of the columns of A; the coefficients of the linear combination are the entries of x. The row vector y^TA is a linear combination of the rows of A; the coefficients of the linear combination are the entries of y.
b) If b_j is the jth column of B and a_i^T is the ith row of A, then the jth column of AB is Ab_j and the ith row of AB is a_i^TB.

To paraphrase, in the matrix product AB, *left multiplication by A multiplies the columns of B* and *right multiplication by B multiplies the rows of A*. See (0.9.1) for an important special case of this observation when one of the factors is a diagonal matrix.

Suppose $A \in M_{m,p}(\mathbf{F})$ and $B \in M_{n,q}$. Let a_k be the kth column of A and let b_k be the kth column of B. Then

c) If $m = n$ then $A^T B = \left[a_i^T b_j\right]$: the i, j entry of $A^T B$ is the scalar $a_i^T b_j$.
d) If $p = q$, then $AB^T = \sum_{k=1}^n a_k b_k^T$: each summand is an m-by-n matrix, the *outer product* of a_k and b_k.

0.2.7 Column space and row space of a matrix The range of $A \in M_{m,n}(\mathbf{F})$ is also called its *column space* because Ax is a linear combination of the columns of A for any $x \in \mathbf{F}^n$ (the entries of x are the coefficients in the linear combination); the set $\{y^T A : y \in \mathbf{F}^m\}$ is the *row space* of A. If the column space of $A \in M_{m,n}(\mathbf{F})$ is contained in the column space of $B \in M_{m,k}(\mathbf{F})$, then there is some $X \in M_{k,n}(\mathbf{F})$ such that $A = BX$ (and conversely); the entries in column j of X tell how to express column j of A as a linear combination of the columns of B.

If $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{m,q}(\mathbf{F})$, then

$$\operatorname{range} A + \operatorname{range} B = \operatorname{range} \left[\begin{array}{cc} A & B \end{array}\right] \tag{0.2.7.1}$$

if $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{p,n}(\mathbf{F})$, then

$$\operatorname{nullspace} A \cap \operatorname{nullspace} B = \operatorname{nullspace} \left[\begin{array}{c} A \\ B \end{array}\right] \tag{0.2.7.2}$$

0.3 Determinants

Often in mathematics it is useful to summarize a multivariate phenomenon with a single number, and the determinant function is an example of this. Its domain is $M_n(\mathbf{F})$ (square matrices only), and it may be presented in several different ways. We denote the determinant of $A \in M_n(\mathbf{F})$ by $\det A$.

0.3.1 Laplace expansion by minors along a row or column. The determinant may be defined inductively for $A = [a_{ij}] \in M_n(\mathbf{F})$ in the following way. Assume that the determinant is defined over $M_{n-1}(\mathbf{F})$ and let $A_{ij} \in M_{n-1}(\mathbf{F})$ denote the submatrix of $A \in M_n(\mathbf{F})$ obtained by deleting row i and column j of A. Then for any $i, j \in \{1, \ldots, n\}$ we have

$$\det A = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det A_{ik} = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det A_{kj} \tag{0.3.1.1}$$

The first sum is the *Laplace expansion by minors along row* i; the second sum is the *Laplace expansion by minors along column* j. This inductive presentation begins by defining the determinant of a 1-by-1 matrix to be the value of the single entry. Thus,

$$\begin{aligned}
\det\,[\,a_{11}] &= a_{11} \\
\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} &= a_{11}a_{22} - a_{12}a_{21} \\
\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\
&\quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}
\end{aligned}$$

and so on. Notice that $\det A^T = \det A$, $\det A^* = \overline{\det\ A}$ if $A \in M_n(\mathbf{C})$, and $\det I = 1$.

0.3.2 Alternating sum. Consistent with the low-dimensional examples in (0.3.1), for $A = [a_{ij}] \in M_n(\mathbf{F})$ we have the alternative presentation

$$\det A = \sum_{\sigma} \operatorname{sgn}\sigma \prod_{i=1}^{n} a_{i\sigma(i)} \tag{0.3.2.1}$$

in which the sum is over all $n!$ permutations of the n items $\{1, \ldots, n\}$ and $\operatorname{sgn}\sigma$, the "sign" or "signum" of a permutation σ, is $+1$ or -1 according to whether the minimum number of transpositions (pair-wise interchanges) necessary to achieve it starting from $\{1, 2, \ldots, n\}$ is even or odd. Thus, each product $a_{1\sigma(1)}a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ enters into the sum defining the determinant with a $+$ sign if σ is even or a $-$ sign if σ is odd.

If $\operatorname{sgn}\sigma$ in (0.3.2.1) is replaced by certain other functions of σ, one obtains *generalized matrix functions* in place of $\det A$. For example, the *permanent* of A, denoted by $\operatorname{per} A$, is obtained by replacing $\operatorname{sgn}\sigma$ by the function that is identically $+1$.

0.3.3 Elementary row and column operations. Three simple and fundamental operations on rows or columns, called *elementary row and column operations* can be used to transform a matrix (square or not) into a simple form that facilitates such tasks as solving linear equations, determining rank, and calculating determinants and inverses of square matrices. We focus on *row operations*, which are implemented by matrices that act on the left. *Column operations* are defined and used in a similar fashion; the matrices that implement them act on the right.

Type 1: Interchange of two rows

For $i \neq j$, interchange of rows i and j of A results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

The two off-diagonal 1's are in the i, j and j, i positions, the two diagonal 0's are in positions i, i and j, j, and all unspecified entries are 0.

Type 2: Multiplication of a row by a nonzero scalar

Multiplication of row i of A by a nonzero scalar c results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & c & & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

The i, i entry is c, all other main diagonal entries are 1, and all unspecified entries are 0.

Type 3: Addition of a scalar multiple of one row to another row

For $i \neq j$, addition of c times row i of A to row j of A results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 1 & \\ & & c & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

The j, i entry is c, all main diagonal entries are 1, and all unspecified entries are 0. The displayed matrix illustrates the case in which $j > i$.

The matrices of each of the three elementary row (or column) operations are just the result of the applying the respective operation to the identity matrix I (on the left for a row operation; on the right for a column operation). The effect of a Type 1 operation on the determinant is to multiply it by -1; the effect of a Type 2 operation is to multiply it by the nonzero scalar c; a Type 3 elementary operation does not change the determinant. The determinant of a square matrix with a zero row is zero. The determinant of a square matrix is zero if and only if some subset of the rows of the matrix is dependent.

0.3.4 Row-reduced echelon form. To each $A \in M_{m,n}(\mathbf{F})$ there corresponds a canonical form in $M_{m,n}(\mathbf{F})$, the *row-reduced echelon form* (RREF) of A, which may be attained by a sequence of elementary row operations. Many matrices have the same RREF, but each matrix has only one RREF regardless of the sequence of elementary operations used to attain it. The defining specifications of the RREF are:

(a) Each nonzero row has 1 as its first nonzero entry;
(b) All other entries in the column of such a leading 1 are 0;
(c) Any rows consisting entirely of 0's occur at the bottom of the matrix; and
(d) The leading 1's occur in a stairstep pattern, left to right; that is, a leading 1 in a lower row must occur to the right of its counterpart above it.

For example,

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

is in RREF.

If $R \in M_{m,n}(\mathbf{F})$ is the RREF of A, then $R = EA$, in which the nonsingular matrix $E \in M_m(\mathbf{F})$ is a product of Type 1, Type 2, and Type 3 elementary matrices corresponding to the sequence of elementary row operations performed to reduce A to RREF.

The determinant of $A \in M_n(\mathbf{F})$ is nonzero if and only if its RREF is I_n. The value of $\det A$ may be calculated by recording the effects upon the determinant of each of the elementary operations that lead to the RREF.

For the system of linear equations $Ax = b$ with $A \in M_{m,n}(\mathbf{F})$ and $b \in \mathbf{F}^m$ given and $x \in \mathbf{F}^n$ unknown, the set of solutions is unchanged if the same sequence of elementary row operations is performed on both A and b. The solutions of $Ax = b$ are revealed by inspection of the RREF of $[A\ b]$. Since the RREF is unique, two systems of linear equations $A_i x = b_i$, $i = 1, 2$, have the same set of solutions if and only if the two augmented matrices $[A_i\ b_i]$, $i = 1, 2$, have the same RREF.

0.3.5 Multiplicativity. A key property of the determinant function is that it is multiplicative: For $A, B \in M_n(\mathbf{F})$

$$\det AB = \det A \ \det B$$

This may be proved using elementary operations that row-reduce both A and B.

0.3.6 Functional characterization of the determinant. If we think of the determinant as a function of each row (or column) of a matrix separately with the others fixed, the Laplace expansion (0.3.1.1) reveals that the determinant is a linear function of the entries in any one given row (column). We summarize this property by saying that the function $A \to \det A$ is *multilinear* in the rows (columns) of A.

The determinant function $A \to \det \mathbf{A}$ is the unique function $f : M_n(\mathbf{F}) \to \mathbf{F}$ that is

(a) Multilinear in the rows of its argument;
(b) Alternating: any Type 1 operation on A changes the sign of $f(A)$; and
(c) Normalized: $f(I) = 1$.

The permanent function is also multilinear (as are other generalized matrix functions) and it is normalized, but it is not alternating.

0.4 Rank

0.4.1 Definition. If $A \in M_{m,n}(\mathbf{F})$, $\operatorname{rank} A$ is the largest number of columns of A that constitute an independent set. There can be different sets of independent columns whose cardinality (number of elements) equals the rank. It is a remarkable fact that $\operatorname{rank} A^T = \operatorname{rank} A$. Therefore, an equivalent definition of rank is the largest number of rows of A that constitute an independent set. Thus, row rank = column rank.

0.4.2 Rank and linear systems. Let $A \in M_{m,n}(\mathbf{F})$ and $b \in \mathbf{F}^n$ be given. The linear system $Ax = b$ may have no solution, exactly one solution, or infinitely many solutions; these are the only possibilities. If there is at least one solution, the system is said to be *consistent*; if there is no solution, the system is said to be *inconsistent*. The linear system $Ax = b$ is consistent if and only if $\operatorname{rank}[A \ b] = \operatorname{rank} A$. The matrix $[A \ b] \in M_{m,n+1}(\mathbf{F})$ is called the *augmented matrix*. To say that the augmented matrix and the *coefficient matrix* A of a linear system have the same rank is just to say that b is a linear combination of the columns of A. In this case, appending b to the columns of A does not increase the rank. A solution of the linear system $Ax = b$ is a vector x whose entries are the coefficients in a representation of b as a linear combination of the columns of A.

0.4.3 RREF and rank. Elementary operations do not change the rank of a matrix, and thus $\operatorname{rank} A$ is the same as the rank of the RREF of A, which is just the number of nonzero rows in the RREF. As a practical matter, however, numerical calculation of the rank by calculation of the RREF is unwise. Round-off errors in intermediate numerical calculations can make zero rows of the RREF appear to be nonzero, thereby affecting perception of the rank.

0.4.4 Characterizations of rank. The following statements about a given matrix $A \in M_{m,n}(\mathbf{F})$ are all equivalent; each can be useful in a different context.

(a) $\operatorname{rank} A = k$;
(b) There exist k, and no more than k, rows of A that constitute a linearly independent set;
(c) There exist k, and no more than k, columns of A that constitute a linearly independent set;
(d) There is a k-by-k submatrix of A with nonzero determinant, but all $(k+1)$-by-$(k+1)$ submatrices of A have determinant 0;
(e) $\dim(\operatorname{range} A) = k$;
(f) There is a set of k, but no more than k, linearly independent vectors b such that the linear system $Ax = b$ is consistent;
(g) $k = n - \dim(\operatorname{nullspace} A)$ (the *rank-nullity theorem*);
(h) $k = \min\{p : A = XY^T$ for some $X \in M_{m,p}(\mathbf{F}), Y \in M_{n,p}(\mathbf{F})\}$;
(i) $k = \min\{p : A = x_1 y_1^T + \cdots + x_p y_p^T\}$ for some $x_1, \ldots, x_p \in \mathbf{F}^m, y_1, \ldots, y_p \in \mathbf{F}^n\}$

0.4.5 Rank inequalities. Some fundamental inequalities involving the rank are:

(a) If $A \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} A \leq \min\{m, n\}$.
(b) If one or more rows and/or columns are deleted from a matrix, the rank of the resulting submatrix is not greater than the rank of the original matrix.

(c) Sylvester's inequality: If $A \in M_{m,k}(\mathbf{F})$ and $B \in M_{k,n}(\mathbf{F})$, then

$$(\operatorname{rank} A + \operatorname{rank} B) - k \leq \operatorname{rank} AB \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}$$

(d) The rank-sum inequality: If $A, B \in M_{m,n}(\mathbf{F})$, then

$$|\operatorname{rank} A - \operatorname{rank} B| \leq \operatorname{rank}(A+B) \leq \operatorname{rank} A + \operatorname{rank} B \quad (0.4.5.1)$$

with equality in the second inequality if and only if $(\operatorname{range} A) \cap (\operatorname{range} B) = \{0\}$ and $(\operatorname{range} A^T) \cap (\operatorname{range} B^T) = \{0\}$. If $\operatorname{rank} B = 1$ then

$$|\operatorname{rank}(A+B) - \operatorname{rank} A| \leq 1 \qquad (0.4.5.2)$$

in particular, changing one entry of a matrix can change its rank by at most 1.

(e) Frobenius's inequality: If $A \in M_{m,k}(\mathbf{F}), B \in M_{k,p}(\mathbf{F})$, and $C \in M_{p,n}(\mathbf{F})$, then

$$\operatorname{rank} AB + \operatorname{rank} BC \leq \operatorname{rank} B + \operatorname{rank} ABC$$

with equality if and only if there are matrices X and Y such that $B = BCX + YAB$.

0.4.6 Rank equalities.

(a) If $A \in M_{m,n}(\mathbf{C})$, then rank $A^* =$ rank $A^T =$ rank $\bar{A} =$ rank A.
(b) If $A \in M_m(\mathbf{F})$ and $C \in M_n(\mathbf{F})$ are nonsingular and $B \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} AB = \operatorname{rank} B = \operatorname{rank} BC = \operatorname{rank} ABC$; that is, left or right multiplication by a nonsingular matrix leaves rank unchanged.
(c) If $A, B \in M_{m,n}(\mathbf{F})$, then rank $A =$ rank B if and only if there exist a nonsingular $X \in M_m(\mathbf{F})$ and a nonsingular $Y \in M_n(\mathbf{F})$ such that $B = XAY$.
(d) If $A \in M_{m,n}(\mathbf{C})$, then rank $A^*A =$ rank A.
(e) *Full-rank factorization*: If $A \in M_{m,n}(\mathbf{F})$, then A has rank k if and only if $A = XY^T$ for some $X \in M_{m,k}(\mathbf{F})$ and $Y \in M_{n,k}(\mathbf{F})$ that each have independent columns. In particular, A has rank 1 if and only if $A = xy^T$ for some nonzero vectors $x \in \mathbf{F}^m$ and $y \in \mathbf{F}^n$. The equivalent factorization $A = XBY^T$ for some nonsingular $B \in M_k(\mathbf{F})$ can also be useful.
(f) Let $A \in M_{m,n}(\mathbf{F})$. If $X \in M_{n,k}(\mathbf{F})$ and $Y \in M_{m,k}(\mathbf{F})$, and if $W = Y^TAX$ is nonsingular, then

$$\operatorname{rank}(A - AXW^{-1}Y^TA) = \operatorname{rank} A - \operatorname{rank} AXW^{-1}Y^TA \quad (0.4.6.1)$$

When $k = 1$, this is *Wedderburn's rank-one reduction formula*: If $x \in \mathbf{F}^n$ and $y \in \mathbf{F}^m$, and if $\omega = y^T Ax \neq 0$, then

$$\operatorname{rank}\left(A - \omega^{-1} Axy^T A\right) = \operatorname{rank} A - 1 \tag{0.4.6.2}$$

Conversely, if $\sigma \in \mathbf{F}$, $u \in \mathbf{F}^n$, $v \in \mathbf{F}^m$, and $\operatorname{rank}\left(A - \sigma uv^T\right) < \operatorname{rank} A$, then $\operatorname{rank}\left(A - \sigma uv^T\right) = \operatorname{rank} A - 1$ and there are $x \in \mathbf{F}^n$ and $y \in \mathbf{F}^m$ such that $u = Ax$, $v = A^T y$, $y^T Ax \neq 0$, and $\sigma = (y^T Ax)^{-1}$.

0.5 Nonsingularity

A linear transformation or matrix is said to be *nonsingular* if it produces the output 0 only for the input 0. Otherwise, it is *singular*. If $A \in M_{m,n}(\mathbf{F})$ and $m < n$, then *A* is necessarily singular. An $A \in M_n(\mathbf{F})$ is *invertible* if there is a matrix $A^{-1} \in M_n(\mathbf{F})$ (the *inverse* of *A)* such that $A^{-1}A = I$. If $A \in M_n$ and $A^{-1}A = I$, then $AA^{-1} = I$, that is, A^{-1} is a *left inverse* if and only if it is a *right inverse*; A^{-1} is unique whenever it exists.

It is useful to be able to call on several different criteria for a square matrix to be nonsingular. The following are equivalent for a given $A \in M_n(\mathbf{F})$:

(a) *A* is nonsingular;
(b) A^{-1} exists;
(c) rank $A = n$;
(d) The rows of *A* are linearly independent;
(e) The columns of *A* are linearly independent;
(f) det $A \neq 0$;
(g) The dimension of the range of *A* is *n*;
(h) The dimension of the null space of *A* is 0;
(i) $Ax = b$ is consistent for each $b \in \mathbf{F}^n$;
(j) If $Ax = b$ is consistent, then the solution is unique;
(k) $Ax = b$ has a unique solution for each $b \in \mathbf{F}^n$;
(l) The only solution to $Ax = 0$ is $x = 0$; and
(m) 0 is not an eigenvalue of *A* (see Chapter 1).

The conditions (g) and (h) are equivalent for a linear transformation $T : V \to V$ on a finite dimensional vector space V: $Tx = y$ has a solution x for every $y \in V$ if and only if the only x such that $Tx = 0$ is $x = 0$ if and only if $Tx = y$ has a *unique* solution x for every $y \in V$.

The nonsingular matrices in $M_n(\mathbf{F})$ form a group, the *general linear group* , often denoted $GL(n, \mathbf{F})$.

If $A \in M_n(\mathbf{F})$ is nonsingular, then $((A^{-1})^T A^T)^T = A(A^{-1}) = I$, so

$(A^{-1})^T A^T = I$, which means that $(A^{-1})^T = (A^T)^{-1}$. It is often convenient to write either $(A^{-1})^T$ or $(A^T)^{-1}$ as A^{-T}. If $A \in M_n(\mathbf{C})$ is nonsingular, then $(A^{-1})^* = (A^*)^{-1}$ and we may safely write either as A^{-*}.

0.6 The Euclidean inner product and norm

0.6.1 Definitions. The scalar $\langle x, y\rangle \equiv y^* x$ is the *Euclidean inner product* (*standard inner product, usual inner product, scalar product*) of $x, y \in \mathbf{C}^n$. The *Euclidean norm* (*usual norm, Euclidean length*) on $\mathbf{C}^n$ is the real-valued function $\|x\|_2 = \langle x, x\rangle^{1/2} = (x^* x)^{1/2}$; two important properties of this function are that $\|x\|_2 > 0$ for all nonzero $x \in \mathbf{C}^n$ and $\|\alpha x\|_2 = |\alpha| \, \|x\|_2$ for all $x \in \mathbf{C}^n$ and all $\alpha \in \mathbf{C}$. The function $\langle \bullet, \bullet\rangle : \mathbf{C}^n \times \mathbf{C}^n \to \mathbf{C}$ is linear in the first argument and *conjugate linear* in the second, that is, $\langle \alpha x_1 + \beta x_2, y\rangle = \alpha\langle x_1, y\rangle + \beta\langle x_2, y\rangle$ and $\langle x, \alpha y_1 + \beta y_2\rangle = \bar{\alpha}\langle x, y_1\rangle + \bar{\beta}\langle x, y_2\rangle$ for all $\alpha, \beta \in \mathbf{C}$ and $y_1, y_2 \in \mathbf{C}^n$. A function $f : \mathbf{C}^n \times \mathbf{C}^n \to \mathbf{C}$ with these two properties is said to be *sesquilinear*; f is an *inner product* if it is sesquilinear and $f(x, x) > 0$ for every nonzero $x \in \mathbf{C}^n$.

0.6.2 Orthogonality and orthonormality. Two vectors $x, y \in \mathbf{C}^n$ are *orthogonal* if $\langle x, y\rangle = 0$. In $\mathbf{R}^2$ and $\mathbf{R}^3$, "orthogonal" has the conventional geometric interpretation of "perpendicular". A set of vectors $S \subset \mathbf{C}^n$ is said to be *orthogonal* if each pair of vectors $x, y \in S$ is orthogonal. A set of orthogonal vectors, none of which is the zero vector, is linearly independent. A vector whose Euclidean norm is 1 is said to be *normalized* (a *unit vector*). For any nonzero $x \in \mathbf{C}^n$, $x / \|x\|_2$ is a unit vector. An orthogonal set of vectors is said to be an *orthonormal* set if each of its elements is a unit vector.

0.6.3 The Cauchy–Schwarz inequality. The *Cauchy–Schwarz inequality* states that

$$|\langle x, y\rangle| \leq \|x\|_2 \, \|y\|_2$$

for all $x, y \in \mathbf{C}^n$, with equality if and only if one of the vectors is a scalar multiple of the other. The *angle* θ between two nonzero vectors $x, y \in \mathbf{C}^n$ is defined by

$$\cos\theta = \frac{|\langle x, y\rangle|}{\|x\|_2 \, \|y\|_2}, \quad 0 \leq \theta \leq \frac{\pi}{2}$$

0.6.4 Gram–Schmidt orthonormalization. Any finite independent set $S \subset \mathbf{C}^n$ may be replaced by an orthonormal set with the same span. This replacement may be carried out in many ways, but there is a systematic way to do so that has a useful special property. Present S as an ordered list: $x_1, \ldots, x_n$. The *Gram–Schmidt process* produces a list of vectors $z_1, \ldots, z_n$ such that $\operatorname{span}\{z_1, \ldots, z_k\} = \operatorname{span}\{x_1, \ldots, x_k\}$ for each $k = 1, \ldots, n$ and the set $\{z_1, \ldots, z_n\}$ is orthonormal. The vectors z_i may be calculated in turn as follows: Let $y_1 = x_1$ and normalize it: $z_1 = y_1 / \|y_1\|_2$. Let $y_2 = x_2 - \langle x_2, z_1 \rangle z_1$ (y_2 is orthogonal to z_1) and normalize it: $z_2 = y_2 / \|y_2\|_2$. Once $z_1, \ldots, z_{k-1}$ have been determined, the vector

$$y_k = x_k - \langle x_k, z_{k-1} \rangle z_{k-1} - \langle x_k, z_{k-2} \rangle z_{k-2} - \cdots - \langle x_k, z_1 \rangle z_1$$

is orthogonal to $z_1, \ldots, z_{k-1}$; normalize it: $z_k = y_k / \|y_k\|_2$. Continue until $k = n$. If we denote $Z = [z_1 \ z_2 \ \ldots \ z_n]$ and $X = [x_1 \ x_2 \ \ldots \ x_n]$, the Gram-Schmidt process gives a factorization $Z = XR$, in which the square matrix *R* $= [r_{ij}]$ is nonsingular and upper triangular; that is, $r_{ij} = 0$ whenever $i > j$.

The Gram–Schmidt process may be applied to any finite list of vectors, independent or not. If $\{x_1, \ldots, x_n\}$ is not independent, the Gram-Schmidt process produces a vector $y_k = 0$ for the least value of *k* for which $\{x_1, \ldots, x_k\}$ is dependent, that is, the least value of k such that x_k is a linear combination of $x_1, \ldots, x_{k-1}$.

0.6.5 Orthonormal bases. An orthonormal set of vectors does not contain the zero vector and is linearly independent. An *orthonormal basis* is a basis whose elements constitute an orthonormal set. Since any finite ordered basis may be transformed with the Gram-Schmidt process to an orthonormal basis, any finite-dimensional real or complex vector space has an orthonormal basis, and any orthonormal set may be extended to an orthonormal basis. Such a basis is pleasant to work with, since the cross terms in inner product calculations all vanish.

0.6.6 Orthogonal complements. Given any subset $S \subset \mathbf{C}^n$, the *orthogonal complement* of S is the set $S^\perp \equiv \{x \in \mathbf{C}^n : x^* y = 0 \text{ for all } y \in S\}$. Even if S is not a subspace, $S^\perp$ is always a subspace. We have $(S^\perp)^\perp = \operatorname{span} S$, and $(S^\perp)^\perp = S$ if S is a subspace. It is always the case that $\dim S^\perp + \dim (S^\perp)^\perp = n$. If S_1 and S_2 are subspaces, then $(S_1 + S_2)^\perp = S_1^\perp \cap S_2^\perp$.

In the context of the linear system $Ax = b$ with $A \in M_{m,n}$, $\operatorname{range} A$ is the orthogonal complement of $\operatorname{nullspace} A^*$; that is, $Ax = b$ has solution (not necessarily unique) if and only if $b^* z = 0$ for all $z \in \mathbf{C}^m$ such that $A^* z = 0$.

If $A \in M_{m,n}$ and $B \in M_{m,q}$, if $X \in M_{m,r}$ and $Y \in M_{m,s}$, and if $\operatorname{range} X = \operatorname{nullspace} A^*$ and $\operatorname{range} Y = \operatorname{nullspace} B^*$, then we have the following companion to (0.2.7.1) and (0.2.7.2)

$$\operatorname{range} A \cap \operatorname{range} B = \operatorname{nullspace} \begin{bmatrix} X^* \\ Y^* \end{bmatrix} \tag{0.6.6.1}$$

0.7 Partitioned sets and matrices

A *partition* of set $\mathcal{S}$ is a collection of subsets of $\mathcal{S}$ such that each element of $\mathcal{S}$ is a member of one and only one of the subsets. For example, a partition of the set $\{1, 2, \dots, n\}$ is a collection of subsets $\alpha_1, \dots, \alpha_t$ (called *index sets*) such that each integer between 1 and n is in one and only one of the index sets. A *sequential partition* of $\{1, 2, \dots, n\}$ is a partition in which the index sets have the special form $\alpha_1 = \{1, \dots, i_1\}, \alpha_2 = \{i_1 + 1, \dots, i_2\}, \dots, \alpha_t = \{i_{t-1} + 1, \dots, n\}$.

A *partition* of a matrix is a decomposition of the matrix into submatrices such that each entry of the original matrix is in one and only one of the submatrices. Partitioning of matrices is often a convenient device for perception of useful structure. For example, partitioning $B = [b_1\ b_2\ \dots\ b_n] \in M_n(\mathbf{F})$ according to its columns reveals the presentation $AB = [Ab_1\ Ab_2\ \dots\ Ab_n]$ of the matrix product, partitioned according to the columns of AB.

0.7.1 Submatrices. Let $A \in M_{m,n}(\mathbf{F})$. For index sets $\alpha \subseteq \{1, \dots, m\}$ and $\beta \subseteq \{1, \dots, n\}$, we denote the (sub)matrix that lies in the rows of A indexed by α and the columns indexed by β as $A[\alpha, \beta]$. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} [\{1,3\}, \{1,2,3\}] = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

If $m = n$ and $\beta = \alpha$, the submatrix $A[\alpha] := A[\alpha, \alpha]$ is a *principal submatrix* of A. An n-by-n matrix has $\binom{n}{k}$ distinct principal submatrices of size k.

Often it is convenient to indicate a submatrix or principal submatrix via deletion, rather than inclusion, of rows or columns. This may be accomplished by complementing the index sets. Let $\alpha^c = \{1, \dots, m\} \backslash \alpha$ and $\beta^c = \{1, \dots, n\} \backslash \beta$ denote the index sets complementary to α and β, respectively. Then $A[\alpha^c, \beta^c]$ is the submatrix obtained by *deleting* the rows indexed by α and the columns indexed by β. For example, the submatrix $A[\alpha, \emptyset^c]$ contains the rows of A indexed by α; $A[\emptyset^c, \beta]$ contains the columns of A indexed by β.

The determinant of a square submatrix of *A* is called a *minor* of *A*. If the submatrix is a principal submatrix, then the minor is a *principal minor*. A signed minor, such as those appearing in the Laplace expansion (0.3.1.1) $[(-1)^{i+j} \det A_{ij}]$ is called a *cofactor* of *A*. By convention, the empty principal minor is 1; that is, $\det A[\emptyset] = 1$. The *size* of a cofactor or minor is the number of rows of the underlying square submatrix.

0.7.2 Partitions, block matrices, and multiplication. If $\alpha_1, \ldots, \alpha_t$ constitute a partition of $\{1, \ldots, m\}$ and $\beta_1, \ldots, \beta_s$ constitute a partition of $\{1, \ldots, n\}$, then the matrices $A\left[\alpha_i, \beta_j\right]$ form a partition of the matrix $A \in M_{m,n}(\mathbf{F})$, $1 \leq i \leq t$, $1 \leq j \leq s$. If $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{n,p}(\mathbf{F})$ are partitioned so that the two partitions of $\{1, \ldots, n\}$ coincide, the two matrix partitions are said to be *conformal*. In this event,

$$(AB)\left[\alpha_i, \gamma_j\right] = \sum_{k=1}^{s} A\left[\alpha_i, \beta_k\right] B\left[\beta_k, \gamma_j\right] \qquad (0.2.7.1)$$

in which the respective collections of submatrices $A[\alpha_i, \beta_k]$ and $B[\beta_k, \gamma_j]$ are conformal partitions of *A* and *B*, respectively. The left-hand side of (0.2.7.1) is a submatrix of the product *AB* (calculated in the usual way), and each summand on the right-hand side is a standard matrix product. Thus, multiplication of conformally partitioned matrices mimics usual matrix multiplication. The sum of two partitioned matrices $A, B \in M_{m,n}(\mathbf{F})$ of the same size has a similarly pleasant representation if the partitions of their rows (respectively, of their the columns) are the same:

$$(A + B)\left[\alpha_i, \beta_j\right] = A\left[\alpha_i, \beta_j\right] + B\left[\alpha_i, \beta_j\right]$$

If a matrix is partitioned by sequential partitions of its rows and columns, the resulting partitioned matrix is called a *block matrix*. For example, if the rows and columns of $A \in M_n(\mathbf{F})$ are partitioned by the same sequential partition $\alpha_1 = \{1, \ldots, k\}, \alpha_2 = \{k+1, \ldots, n\}$, the resulting block matrix is

$$A = \begin{bmatrix} A[\alpha_1, \alpha_1] & A[\alpha_1, \alpha_2] \\ A[\alpha_2, \alpha_1] & A[\alpha_2, \alpha_2] \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

in which the *blocks* are $A_{ij} = A[\alpha_i, \alpha_j]$. Computations with block matrices are employed throughout the book; 2-by-2 block matrices are the most important and useful.

0.7.3 The inverse of a partitioned matrix. It can be useful to know the corresponding blocks in the inverse of a partitioned nonsingular matrix *A*, that is, to present the inverse of a partitioned matrix in conformally partitioned form. This may be done in a variety of apparently different, but equivalent, ways—assuming that certain submatrices of $A \in M_n(\mathbf{F})$ and A^{-1} are also nonsingular. For simplicity, let A be partitioned as a 2-by-2 block matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with $A_{ii} \in M_{n_i}(\mathbf{F}), i = 1, 2$ and $n_1 + n_2 = n$. A useful expression for the correspondingly partitioned presentation of A^{-1} is

$$\begin{bmatrix} \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} & A_{11}^{-1}A_{12}\left(A_{21}A_{11}^{-1}A_{12} - A_{22}\right)^{-1} \\ A_{22}^{-1}A_{21}\left(A_{12}A_{22}^{-1}A_{21} - A_{11}\right)^{-1} & \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{bmatrix} \tag{0.7.3.1}$$

assuming that all the relevant inverses exist. This expression for A^{-1} may be verified by doing a partitioned multiplication by A and then simplifying. In general index set notation, we may write

$$A^{-1}[\alpha] = \Big(A[\alpha] - A[\alpha, \alpha^c]\, A[\alpha^c]^{-1}\, A[\alpha^c, \alpha]\Big)^{-1}$$

and

$$\begin{aligned} A^{-1}[\alpha, \alpha^c] &= A[\alpha]^{-1}\, A[\alpha, \alpha^c]\Big(A[\alpha^c, \alpha]\, A[\alpha]^{-1}\, A[\alpha, \alpha^c] - A[\alpha^c]\Big)^{-1} \\ &= \Big(A[\alpha, \alpha^c]\, A[\alpha^c]^{-1}\, A[\alpha^c, \alpha] - A[\alpha]\Big)^{-1} A[\alpha, \alpha^c]\, A[\alpha^c]^{-1} \end{aligned}$$

again assuming that the relevant inverses exist. There is an intimate relationship between these representations and the Schur complement; see (0.8.5). Notice that $A^{-1}[\alpha]$ is a submatrix of A^{-1}, while $A[\alpha]^{-1}$ is the inverse of a submatrix of A; these two objects are not, in general, the same.

0.7.4 The Sherman-Morrison-Woodbury formula. Suppose a nonsingular matrix $A \in M_n(\mathbf{F})$ has a known inverse A^{-1} and consider $B = A + XRY$, in which X is n-by-r, Y is r-by-n, and R is r-by-r and nonsingular. If B and $R^{-1} + YA^{-1}X$ are nonsingular, then

$$B^{-1} = A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1} \tag{0.7.4.1}$$

If r is much smaller than n, then R and $R^{-1} + YA^{-1}X$ may be much easier to invert than B. For example, if $x, y \in \mathbf{F}^n$ are nonzero vectors, $X = x, Y = y^T$, $y^TA^{-1}x \neq -1$, and R = [1], then (0.7.4.1) becomes a formula for the inverse of a rank-1 adjustment to A:

$$\left(A + xy^T\right)^{-1} = A^{-1} - \left(1 + y^TA^{-1}x\right)^{-1} A^{-1}xy^TA^{-1} \tag{0.7.4.2}$$

In particular, if $B = I + xy^T$ for $x, y \in \mathbf{F}^n$ and $y^T x \neq -1$, then $B^{-1} = I - (1 + y^T x)^{-1} xy^T$.

0.7.5 Complementary Nullities. Suppose $A \in M_n(\mathbf{F})$ is nonsingular, let α and β be nonempty subsets of $\{1, \ldots, n\}$, and write $|\alpha| = r$ and $|\beta| = s$ for the cardinalities of α and β. The *law of complementary nullities* is

$$\text{nullity}\,(A\,[\alpha, \beta]) = \text{nullity}\,\left(A^{-1}\,[\beta^c, \alpha^c]\right) \tag{0.7.5.1}$$

which is equivalent to the rank identity

$$\text{rank}\,(A\,[\alpha, \beta]) = \text{rank}\,\left(A^{-1}\,[\beta^c, \alpha^c]\right) + r + s - n \tag{0.7.5.2}$$

Since we can permute rows and columns to place first the r rows indexed by α and the s columns indexed by β, it suffices to consider the presentations

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

in which A_{11} and B_{11}^T are r-by-s and A_{22} and B_{22}^T are $(n - r)$-by-$(n - s)$. Then (0.7.5.1) says that nullity $A_{11} =$ nullity B_{22}.

The underlying principle here is very simple. Suppose the nullity of A_{11} is k. If $k \geq 1$, let the columns of $X \in M_{s,k}(\mathbf{F})$ be a basis for the null space of A_{11}. Since A is nonsingular,

$$A \begin{bmatrix} X \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}X \\ A_{21}X \end{bmatrix} = \begin{bmatrix} 0 \\ A_{21}X \end{bmatrix}$$

has full rank, so $A_{21}X$ has k independent columns. But

$$\begin{bmatrix} B_{12}\,(A_{21}X) \\ B_{22}\,(A_{21}X) \end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\ A_{21}X \end{bmatrix} = A^{-1}A \begin{bmatrix} X \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ 0 \end{bmatrix}$$

so $B_{22}(A_{21}X) = 0$ and hence nullity $B_{22} \geq k =$ nullity A_{11}, a statement that is trivially correct if $k = 0$. A similar argument starting with B_{22} shows that nullity $A_{11} \geq$ nullity B_{22}.

Of course, (0.7.5.1) also tells us that nullity $A_{12} =$ nullity B_{12}, nullity $A_{21} =$ nullity B_{21}, and nullity $A_{22} =$ nullity B_{11}. If $r + s = n$, then rank $A_{11} =$ rank B_{22} and rank $A_{22} =$ rank B_{11}, while if $n = 2r = 2s$, then we also have rank $A_{12} =$ rank B_{12} and rank $A_{21} =$ rank B_{21}. Finally, (0.7.5.2) tells us that the rank of an r-by-s submatrix of an n-by-n nonsingular matrix is at least $r + s - n$.

0.7.6 Rank in a partitioned matrix and rank-principal matrices. Partition $A \in M_n(\mathbf{F})$ as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad A_{11} \in M_r(\mathbf{F}), \mathbf{A}_{22} \in M_{n-r}(\mathbf{F})$$

If A_{11} is nonsingular, then of course $\operatorname{rank}[A_{11}\ A_{12}] = r$ and $\operatorname{rank}[A_{11}^T\ A_{21}^T]^T = r$. Remarkably, the converse is true if all the rank of A is located in its first block row and first block column:

$$\text{If } \operatorname{rank} A = \operatorname{rank}[A_{11}\ A_{12}] = \operatorname{rank} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}, \text{ then } A_{11} \text{ is nonsingular} \tag{0.7.6.1}$$

This follows from (0.4.6(c)): If A_{11} is singular, then $\operatorname{rank} A_{11} = k < r$ and there are nonsingular $S, T \in M_r(\mathbf{F})$ such that

$$SA_{11}T = \begin{bmatrix} I_k & 0 \\ 0 & 0_{r-k} \end{bmatrix}$$

Therefore,

$$\hat{A} = \begin{bmatrix} S & 0 \\ 0 & I_{n-r} \end{bmatrix} A \begin{bmatrix} T & 0 \\ 0 & I_{n-r} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & 0_{r-k} \end{bmatrix} & SA_{12} \\ A_{21}T & A_{22} \end{bmatrix}$$

has rank r, as do its first block row and column. Because the rth row of the first block column of $\hat{A}$ is zero, there must be some column in SA_{12} whose rth entry is not zero, which means that $\hat{A}$ has at least $r+1$ independent columns. This contradicts $\operatorname{rank} \hat{A} = \operatorname{rank} A = r$, so A_{11} must be nonsingular.

Suppose $A \in M_n(\mathbf{F})$ and $\operatorname{rank} A = r$. We say that A is *rank principal* if it has a nonsingular r-by-r *principal* submatrix. It follows from (0.7.6.1) that if there is some index set $\alpha \subseteq \{1, \ldots, n\}$ such that

$$\operatorname{rank} A = \operatorname{rank} A\,[\alpha, \emptyset^c] = \operatorname{rank} A\,[\emptyset^c, \alpha] \tag{0.7.6.2}$$

(that is, if there is some set of r linearly independent rows of A such that the corresponding r columns are linearly independent), then A is rank principal; moreover, $A[\alpha]$ is nonsingular.

If $A \in M_n(\mathbf{F})$ is symmetric or skew-symmetric, or if $A \in M_n(\mathbf{C})$ is Hermitian or skew-Hermitian, then $\operatorname{rank} A\,[\alpha, \emptyset^c] = \operatorname{rank} A\,[\emptyset^c, \alpha]$ for every index set α, so A satisfies (0.7.6.2) and is therefore rank principal.

0.7.7 Commutativity and block diagonal matrices Two matrices $A, B \in M_n(\mathbf{F})$ are said to *commute* if $AB = BA$. Commutativity is not typical, but one important instance is encountered frequently. Suppose $\Lambda = [\Lambda_{ij}]_{i,j=1}^{s} \in M_n(\mathbf{F})$ is a block matrix in which $\Lambda_{ij} = 0$ whenever $i \neq j$; $\lambda_i \in \mathbf{F}$ and $\Lambda_{ii} = \lambda_i I_{n_i}$ for each $i = 1, \ldots, s$; and $\lambda_i \neq \lambda_j$ if $i \neq j$. Partition $B = [B_{ij}]_{i,j=1}^{s} \in M_n(\mathbf{F})$ conformally with Λ. Then $\Lambda B = B\Lambda$ if and only if $\lambda_i B_{ij} = B_{ij}\lambda_j$ for each $i, j = 1, \ldots, s$. These identities are satisfied if and only if $B_{ij} = 0$ whenever $i \neq j$. Thus, Λ *commutes with* B *if and only if* B *is block diagonal conformal with* Λ; see (0.9.2).

0.8 Determinants again

Some additional facts about and identities for the determinant are useful for reference.

0.8.1 Compound matrices. The array of all determinants of submatrices of a given size (that is, minors) of a given matrix $A \in M_{m,n}(\mathbf{F})$ is called a *compound matrix* of A. In particular, the $\binom{m}{k}$-by-$\binom{n}{k}$ matrix whose α, β, entry is det $A[\alpha, \beta]$ is called the *kth compound matrix* of A and is denoted by $C_k(A)$ Here, $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ are index sets of cardinality $k \leq \min\{m, n\}$, usually ordered lexicographically, that is, $\{1, 2, 4\}$ before $\{1, 2, 5\}$ before $\{1, 3, 4\}$ and so on. For example, if

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

then

$$C_2(A) = \begin{bmatrix} \det\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} & \det\begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix} & \det\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix} \\ \det\begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix} & \det\begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix} & \det\begin{bmatrix} 2 & 3 \\ 8 & 9 \end{bmatrix} \\ \det\begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} & \det\begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix} & \det\begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} -3 & -6 & -3 \\ -6 & -12 & -6 \\ -3 & -6 & -3 \end{bmatrix}$$

If $A \in M_{m,k}(\mathbf{F})$ and $B \in M_{k,n}(\mathbf{F})$, then

$$C_r(AB) = C_r(A)C_r(B), \quad r \leq \min\{m, k, n\} \tag{0.8.1.1}$$

Also,

$$C_r(tA) = t^r C_r(A), \quad t \in \mathbf{F}$$
If $I \in M_n$, then $C_k(I) = I \in M_{\binom{n}{k}}$
If $A \in M_n$ is nonsingular, then $C_k(A)^{-1} = C_k(A^{-1})$
If $A \in M_{m,n}(\mathbf{F})$, then $C_k(A^T) = C_k(A)^T$

and

If $A \in M_{m,n}(\mathbf{C})$, then $C_k(A^*) = C_k(A)^*$

If $\Delta \in M_n(\mathbf{F})$ is upper triangular [see (0.9.3)], then so is $C_k(\Delta)$, whose main diagonal entries are the $\binom{n}{k}$ possible products of k entries chosen from the main diagonal of Δ.

0.8.2 The adjugate and the inverse. If $A \in M_n(\mathbf{F})$ and $n \geq 2$, the transposed matrix of cofactors of A

$$\operatorname{adj} A \equiv \left[(-1)^{i+j} \det\left[A\left\{j\right\}^c, \left\{i\right\}^c\right]\right] \tag{0.8.2.0}$$

is called the *adjugate* of A; it is also called the *classical adjoint* of A. A calculation using the Laplace expansion for the determinant reveals the basic property of the adjugate

$$(\operatorname{adj} A)\, A = A\, (\operatorname{adj} A) = (\det A)\, I \tag{0.8.2.1}$$

Thus, $\operatorname{adj} A$ is nonsingular if and only if A is nonsingular, and $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

If A is nonsingular, then $A^{-1} = (\operatorname{adj} A)/\det A$, and

$$\operatorname{adj} A = (\det A)\, A^{-1} \tag{0.8.2.2}$$

In particular, $\operatorname{adj}(A^{-1}) = A/\det A = (\operatorname{adj} A)^{-1}$.

Suppose A is singular. There are only two possible values for $\operatorname{rank}\operatorname{adj} A$. If $\operatorname{rank} A \leq n-2$, then every square submatrix of A of size $n-1$ has zero determinant, so $\operatorname{adj} A = 0$. If $\operatorname{rank} A = n-1$, then some submatrix of A of size $n-1$ has nonzero determinant, $\operatorname{adj} A \neq 0$, A has $n-1$ independent columns, and $(\operatorname{adj} A)\, A = (\det A)\, I = 0$, so the null space of $\operatorname{adj} A$ has dimension $n-1$ and hence $\operatorname{rank}\operatorname{adj} A = 1$. If $\operatorname{rank}\operatorname{adj} A = 1$ and $\operatorname{adj} A = xy^T$, then (0.8.2.1) ensures that

$$(Ax)y^T = A(\operatorname{adj} A) = 0 = (\operatorname{adj} A)A = x(y^T A),$$

so $Ax = 0$ and $y^T A = 0$. Conversely, if $\operatorname{rank} A = n-1$, $x \neq 0 \neq y$, $Ax = 0$, and $y^T A = 0$, then there is a nonzero scalar α such that $\operatorname{adj} A = \alpha x y^T$.

The function $A \rightarrow \operatorname{adj} A$ is continuous on M_n (each entry of $\operatorname{adj} A$ is a multinomial in the entries of A) and every matrix in M_n is a limit of nonsingular matrices, so properties of the adjugate can be deduced from continuity and properties of the inverse function. For example, if $A, B \in M_n$ are nonsingular, then $\operatorname{adj}(AB) = (\det AB)(AB)^{-1} = (\det A)(\det B)B^{-1}A^{-1} = (\det B)B^{-1}(\det A)A^{-1} = (\operatorname{adj} B)(\operatorname{adj} A)$. Continuity then ensures that

$$\operatorname{adj}(AB) = (\operatorname{adj} B)(\operatorname{adj} A) \text{ for all } A, B \in M_n \tag{0.8.2.3}$$

For any $c \in \mathbf{F}$ and any $A \in M_n(\mathbf{F})$, $\operatorname{adj}(cA) = c^{n-1} \operatorname{adj} A$. In particular, $\operatorname{adj}(cI) = c^{n-1} I$ and $\operatorname{adj} 0 = 0$.

If A is nonsingular, then

$$\begin{aligned} \operatorname{adj}(\operatorname{adj} A) &= \operatorname{adj}((\det A)A^{-1}) = (\det A)^{n-1} \operatorname{adj} A^{-1} \\ &= (\det A)^{n-1} (A/\det A) = (\det A)^{n-2} A \end{aligned}$$

so continuity ensures that

$$\operatorname{adj}(\operatorname{adj} A) = (\det A)^{n-2} A \text{ for all } A \in M_n \tag{0.8.2.4}$$

If $A+B$ is nonsingular, then $A(A+B)^{-1}B = B(A+B)^{-1}A$, so continuity ensures that

$$A \operatorname{adj}(A+B) B = B \operatorname{adj}(A+B) A \text{ for all } A, B \in M_n \tag{0.8.2.5}$$

Let $A, B \in M_n$ and suppose A commutes with B. If A is nonsingular, then $BA^{-1} = A^{-1}ABA^{-1} = A^{-1}BAA^{-1} = A^{-1}B$, so A^{-1} commutes with B. But $BA^{-1} = (\det A)B \operatorname{adj} A$ and $A^{-1}B = (\det A)(\operatorname{adj} A)B$, so $\operatorname{adj} A$ commutes with B. Continuity ensures that $\operatorname{adj} A$ commutes with B whenever A commutes with B, even if A is singular.

If $A = [a_{ij}]$ is upper triangular then $\operatorname{adj} A \equiv [b_{ij}]$ is upper triangular and each $b_{ii} = \prod_{j \neq i} a_{jj}$; if A is diagonal then so is $\operatorname{adj} A$.

The adjugate is the transpose of the gradient of $\det A$:

$$(\operatorname{adj} A) = \left[\frac{\partial}{\partial a_{ij}} \det A \right]^T \tag{0.8.2.6}$$

If A is nonsingular, it follows from (0.8.2.6) that

$$\left[\frac{\partial}{\partial a_{ij}} \det A \right]^T = (\det A)\, A^{-1} \tag{0.8.2.7}$$

If $A \in M_n$ is nonsingular, then $\operatorname{adj} A^T = (\det A^T)A^{-T} = (\det A)A^{-T} =$

$((\det A)A^{-1})^T = (\operatorname{adj} A)^T$. Continuity ensures that

$$\operatorname{adj} A^T = (\operatorname{adj} A)^T \text{ for all } A \in M_n(\mathbf{F}) \tag{0.8.2.8}$$

A similar argument shows that

$$\operatorname{adj} A^* = (\operatorname{adj} A)^* \text{ for all } A \in M_n \tag{0.8.2.9}$$

Let $A = [a_1 \ \dots \ a_n] \in M_n(\mathbf{F})$ be partitioned according to its columns and let $b \in \mathbf{F}^n$. Define

$$(A \underset{i}{\leftarrow} b) = [a_1 \ \dots \ a_{i-1} \ b \ a_{i+1} \ \dots \ a_n]$$

that is, $(A \underset{i}{\leftarrow} b)$ denotes the matrix whose ith column is b and whose remaining columns coincide with those of A. Examination of the Laplace expansion (0.3.1.1) of $\det (A \underset{i}{\leftarrow} b)$ by minors along column i reveals that it is the ith entry of the vector $(\operatorname{adj} A)\, b$, that is

$$[\det (A \underset{i}{\leftarrow} b)]_{i=1}^n = (\operatorname{adj} A)\, b \tag{0.8.2.10}$$

Applying this vector identity to each column of $C = [c_1 \ \dots \ c_n] \in M_n(\mathbf{F})$ gives the matrix identity

$$[\det (A \underset{i}{\leftarrow} c_j)]_{i,j=1}^n = (\operatorname{adj} A)\, C \tag{0.8.2.11}$$

0.8.3 Cramer's rule. Cramer's rule is a useful way to present analytically a particular entry of the solution to $Ax = b$ when $A \in M_n(\mathbf{F})$ is nonsingular. The identity

$$A\, [\det (A \underset{i}{\leftarrow} b)]_{i=1}^n = A\, (\operatorname{adj} A)\, b = (\det A)\, b$$

follows from (0.8.2.9). If $\det A \neq 0$ we obtain Cramer's rule

$$x_i = \frac{\det(A \underset{i}{\leftarrow} b)}{\det\ A}$$

for the ith entry x_i of the solution vector x. Cramer's rule also follows directly from multiplicativity of the determinant. The system $Ax = b$ may be rewritten as

$$A(I \underset{i}{\leftarrow} x) = A \underset{i}{\leftarrow} b$$

and taking determinants of both sides (using multiplicativity) gives

$$(\det\ A)\ \det(I \underset{i}{\leftarrow} x) = \det(A \underset{i}{\leftarrow} b)$$

But $\det(I \underset{i}{\leftarrow} x) = x_i$, and the formula follows.

0.8.4 Minors of the inverse. An important fact, which generalizes the adjugate formula for the inverse of a nonsingular matrix and relates the minors of A^{-1} to those of $A \in M_n(\mathbf{F})$, is *Jacobi's identity*:

$$\det\ A^{-1}\left[\alpha^c, \beta^c\right] = (-1)^{\left(\sum\limits_{i\in\alpha} i + \sum\limits_{j\in\beta} j\right)} \frac{\det\ A\left[\beta, \alpha\right]}{\det A} \tag{0.8.4.1}$$

For principal submatrices, this formula assumes the simple form

$$\det\ A^{-1}(\alpha^c) = \frac{\det\ A\left[\alpha\right]}{\det\ A} \tag{0.8.4.2}$$

0.8.5 Schur complements and determinantal formulae. Let $A = [a_{ij}] \in M_n(\mathbf{F})$ be given and suppose $\alpha \subseteq \{1, \ldots, n\}$ is an index set such that $A[\alpha]$ is nonsingular. An important formula for $\det A$, based upon the 2-partition of A using α and α^c, is

$$\det A = \det A\left[\alpha\right] \det \left(A\left[\alpha^c\right] - A\left[\alpha^c, \alpha\right] A\left[\alpha\right]^{-1} A\left[\alpha, \alpha^c\right]\right) \tag{0.8.5.1}$$

which generalizes the familiar formula for the determinant of a 2-by-2 matrix. The special matrix

$$A/A\left[\alpha\right] \equiv A\left[\alpha^c\right] - A\left[\alpha^c, \alpha\right] A\left[\alpha\right]^{-1} A\left[\alpha, \alpha^c\right] \tag{0.8.5.2}$$

which also appears in the partitioned form for the inverse in (0.7.3.1), is called the *Schur complement of* $A\left[\alpha\right]$ *in* A. When convenient, we take $\alpha = \{1, \ldots, k\}$ and write A as a 2-by-2 block matrix $A = [A_{ij}]$ with $A_{11} = A[\alpha]$, $A_{22} = A[\alpha^c]$, $A_{12} = A[\alpha, \alpha^c]$, and $A_{21} = A[\alpha^c, \alpha]$. The formula (0.8.5.1) may be verified by computing the determinant of both sides of the identity

$$\begin{aligned} &\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \end{aligned} \tag{0.8.5.3}$$

which contains a wealth of information about the Schur complement $S = [s_{ij}] \equiv A/A_{11} = A_{22} - A_{21}A_{11}^{-1}A_{12}$:

(a) The Schur complement S arises (uniquely) in the lower right corner if linear combinations of the first k rows (respectively, columns) of A are added to the last $n - k$ rows (respectively, columns) in such a way as to produce a zero block in the lower left (respectively, upper right) corner; this is *block Gaussian elimination*, and it is (uniquely) possible because A_{11} is nonsingular. Any submatrix of A that includes A_{11} as a principal submatrix has the same determinant before and after the block eliminations that produce the block diagonal form in (0.8.5.3). Thus, for any index set $\beta = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n-k\}$, if we construct the shifted index set $\tilde{\beta} \equiv \{i_1+k, \ldots, i_m+k\}$, then $\det A[\alpha \cup \tilde{\beta}, \alpha \cup \tilde{\gamma}]$ (before) $= \det(A_{11} \oplus S[\beta, \gamma])$ (after), so

$$\det S\left[\beta, \gamma\right] = \det A\left[\alpha \cup \tilde{\beta}, \alpha \cup \tilde{\gamma}\right] / \det A\left[\alpha\right] \tag{0.8.5.4}$$

For example, if $\beta = \{i\}$ and $\gamma = \{j\}$, then with $\alpha = \{1, \ldots, k\}$ we have

$$\begin{aligned} \det S\left[\beta, \gamma\right] &= s_{ij} \qquad\qquad (0.8.5.5) \\ &= \det A\left[\{1, \ldots k, k+i\}, \{1, \ldots, k, k+j\}\right] / \det A_{11} \end{aligned}$$

so all the entries of S are ratios of minors of A.

(b) $\operatorname{rank} A = \operatorname{rank} A_{11} + \operatorname{rank} S \geq \operatorname{rank} A_{11}$, and $\operatorname{rank} A = \operatorname{rank} A_{11}$ if and only if $A_{22} = A_{21} A_{11}^{-1} A_{12}$;

(c) A is nonsingular if and only if S is nonsingular, since $\det A = \det A_{11} \det S$. If A is nonsingular, then $\det S = \det A / \det A_{11}$.

Suppose A is nonsingular. Then inverting both sides of (0.8.5.3) gives a presentation of the inverse different from that in (0.7.3.1):

$$A^{-1} = \left[\begin{array}{cc} A_{11} + A_{11}^{-1} A_{12} S^{-1} A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12} S^{-1} \\ -S^{-1} A_{21} A_{11}^{-1} & S^{-1} \end{array}\right] \tag{0.8.5.6}$$

Among other things, this tells us that $A^{-1}[\{k+1, \ldots, n\}] = S^{-1}$, so

$$\det A^{-1}\left[\{k+1, \ldots, n\}\right] = \det A_{11} / \det A \tag{0.8.5.7}$$

This is a form of Jacobi's identity (0.8.4.1). Another form results from writing the inverse in terms of the adjugate, which gives

$$\det\left((\operatorname{adj} A)\left[\{k+1, \ldots, n\}\right]\right) = (\det A)^{n-k-1} \det A_{11} \tag{0.8.5.8}$$

When α^c consists of a single element, the Schur complement of $A[\alpha]$ in A is a scalar and (0.8.5.1) reduces to the identity

$$\det A = A\left[\alpha^c\right] \det A\left[\alpha\right] - A\left[\alpha^c, \alpha\right] \left(\operatorname{adj} A\left[\alpha\right]\right) A\left[\alpha, \alpha^c\right] \tag{0.8.5.9}$$

which is valid even if $A[\alpha]$ is singular. For example, if $\alpha = \{1, \ldots, n-1\}$, then $\alpha^c = \{n\}$ and A is presented as a *bordered matrix*

$$A = \left[\begin{array}{cc} \tilde{A} & x \\ y^T & a \end{array}\right]$$

with $a \in \mathbf{F}$, $x, y \in \mathbf{F}^{n-1}$, and $\tilde{A} \in M_{n-1}(\mathbf{F})$; (0.8.5.9) is the *Cauchy expansion* of the determinant of a bordered matrix

$$\det \begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix} = a \det \tilde{A} - y^T \left(\operatorname{adj} \tilde{A} \right) x \tag{0.8.5.10}$$

The Cauchy expansion (0.8.5.10) involves signed minors of A of size $n-2$ (the entries of $\operatorname{adj} \tilde{A}$) and a bilinear form in the entries of a row *and* column; the Laplace expansion (0.3.1.1) involves signed minors of A of size $n-1$ and a linear form in the entries of a row *or* column. If $a \neq 0$, we can use the Schur complement of $[a]$ in A to express

$$\det \begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix} = a \det(\tilde{A} - a^{-1} x y^T)$$

Equating the right-hand side of this identity to that of (0.8.5.10) and setting $a = -1$ gives *Cauchy's formula for the determinant of a rank-one perturbation*

$$\det \left(\tilde{A} + x y^T \right) = \det \tilde{A} + y^T \left(\operatorname{adj} \tilde{A} \right) x \tag{0.8.5.11}$$

The uniqueness property of the Schur complement discussed in **(a)** can be used to derive an identity involving a Schur complement within a Schur complement. Suppose the nonsingular k-by-k block A_{11} is partitioned as a 2-by-2 block matrix $A_{11} = [\mathcal{A}_{ij}]$ in which the upper left ℓ-by-ℓ block $\mathcal{A}_{11}$ is nonsingular. Write $A_{21} = [\mathcal{A}_1 \ \mathcal{A}_2]$, in which $\mathcal{A}_1$ is $(n-k)$-by-ℓ, and write $A_{12}^T = [\mathcal{B}_1^T \ \mathcal{B}_2^T]$, in which $\mathcal{B}_1$ is ℓ-by-$(n-k)$; this gives the refined partition

$$A = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1 \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{B}_2 \\ \mathcal{A}_1 & \mathcal{A}_2 & A_{22} \end{bmatrix}$$

Now add linear combinations of the first ℓ rows of A to the next $k-\ell$ rows in order to reduce $\mathcal{A}_{21}$ to a zero block. The result is

$$A' \equiv \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1 \\ 0 & A_{11}/\mathcal{A}_{11} & \mathcal{B}_2' \\ \mathcal{A}_1 & \mathcal{A}_2 & A_{22} \end{bmatrix}$$

where we have identified the resulting 2,2 block of A' as the (necessarily nonsingular) Schur complement of $\mathcal{A}_{11}$ in A_{11}. Now add linear combinations of the first k rows of A' to the last $n-k$ rows to reduce $[\,\mathcal{A}_1 \ \mathcal{A}_2]$ to a zero block. The result is

$$A'' \equiv \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1 \\ 0 & A_{11}/\mathcal{A}_{11} & \mathcal{B}_2' \\ 0 & 0 & A/A_{11} \end{bmatrix}$$

in which we have identified the resulting 3,3 block of A'' as the Schur complement of A_{11} in A. The lower right 2-by-2 block of A'' must be $A/\mathcal{A}_{11}$, the Schur complement of $\mathcal{A}_{11}$ in A. Moreover, the lower right block of $A/\mathcal{A}_{11}$ must be the Schur complement of $A_{11}/\mathcal{A}_{11}$ in $A/\mathcal{A}_{11}$. This observation is the *quotient property of Schur complements*:

$$A/A_{11} = (A/\mathcal{A}_{11}) \,/\, (A_{11}/\mathcal{A}_{11}) \tag{0.8.5.12}$$

If the four blocks A_{ij} in (0.8.5.3) are square and the same size, and if A_{11} commutes with A_{21}, then

$$\begin{aligned} \det A &= \det A_{11} \det S = \det(A_{11}S) \\ &= \det(A_{11}A_{22} - A_{11}A_{21}A_{11}^{-1}A_{12}) = \det(A_{11}A_{22} - A_{21}A_{12}) \end{aligned}$$

If A_{11} commutes with A_{12}, the same conclusion follows from computing $\det A = (\det S)(\det A) = \det(SA_{11})$. By continuity, the identity

$$\det A = \det(A_{11}A_{22} - A_{21}A_{12}) \tag{0.8.5.13}$$

is valid whenever A_{11} commutes with either A_{21} or A_{12}, even if it is singular. If A_{22} commutes with either A_{12} or A_{21}, a similar argument using the Schur complement of A_{22} shows that

$$\det A = \det(A_{11}A_{22} - A_{12}A_{21}) \tag{0.8.5.14}$$

if A_{22} commutes with either A_{21} or A_{12}.

0.8.6 Determinantal identities of Sylvester and Kronecker We consider two consequences of (0.8.5.4). If we set

$$B \equiv [b_{ij}] = [\det A\,[\{1,\ldots k, k+i\}\,,\{1,\ldots,k,k+j\}]]_{i,j=1}^{n-k}$$

then each entry of B is the determinant of a bordered matrix of the form (0.8.5.10): $\tilde{A}$ is A_{11}, x is the jth column of A_{12}, y^T is the ith row of A_{21}, and a is the i,j entry of A_{22}. The identity (0.8.5.5) tells us that $B = (\det A_{11})S$, so

$$\begin{aligned} \det B &= (\det A_{11})^{n-k} \det S \\ &= (\det A_{11})^{n-k} (\det A/\det A_{11}) = (\det A_{11})^{n-k-1} \det A \end{aligned}$$

This observation about B is *Sylvester's identity for bordered determinants*:

$$\det B = (\det A\,[\alpha])^{n-k-1} \det A \tag{0.8.6.1}$$

where $B = [\det A[\alpha \cup \{i\}\,, \alpha \cup \{j\}]]$, and i, j are indices *not* contained in α.

If $A_{22} = 0$, then each entry of B is the determinant of a bordered matrix of the form (0.8.5.10) with $a = 0$. In this case, the Schur complement $A/A_{11} = -A_{21}A_{11}^{-1}A_{12}$ has rank at most k, so the determinant of every $(k+1)$-by-$(k+1)$ submatrix of B is zero; this observation about B is *Kronecker's theorem for bordered determinants*.

0.8.7 The Cauchy–Binet formula. This useful formula can be remembered because of its similarity in appearance to the formula for matrix multiplication. This is no accident, since it is equivalent to multiplicativity of the compound matrix (0.8.1.1). Let $A \in M_{m,k}(\mathbf{F})$, $B \in M_{k,n}(\mathbf{F})$, and $C = AB$. Further, let $1 \leq r \leq \min\{m, k, n\}$, and let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be index sets, each of cardinality *r*. An expression for the α, β minor of *C* is

$$\det C\,[\alpha, \beta] = \sum_{\gamma} \det\ A\,[\alpha, \gamma] \det B\,[\gamma, \beta]$$

in which the sum is taken over all index sets $\gamma \subseteq \{1, \ldots, k\}$ of cardinality *r*.

0.8.8 Relations among minors. Let $A \in M_{m,n}(\mathbf{F})$ be given and let a fixed index set $\alpha \subseteq \{1, \ldots, m\}$ of cardinality *k* be given. The minors $\det\ A\,[\alpha, \omega]$, as $\omega \subseteq \{1, \ldots, n\}$ runs over *ordered* index sets of cardinality *k*, are not algebraically independent since there are more minors than there are distinct entries among the submatrices. Quadratic relations are known among these minors. Let $i_1, i_2, \ldots, i_k \in \{1, \ldots, n\}$ be *k* distinct indices, not necessarily in natural order, and let $A[\alpha; i_1, \ldots, i_k]$ denote the matrix whose rows are indicated by α and whose *j*th column is column i_j of $A[\alpha, \{1, \ldots, n\}]$. The difference between this and our prior notation is that columns might not occur in natural order as in $A(\{1, 3\}; 4, 2)$, whose first column has the 1, 4 and 3, 4 entries of *A*. We then have the relations

$$\begin{aligned} &\det A\,[\alpha; i_1, \ldots, i_k] \det\ A\,[\alpha; j_1, \ldots, j_k] \\ = &\sum_{t=1}^{k} \det\ A\,[\alpha; i_1, \ldots, i_{s-1}, j_t, i_{s+1}, \ldots, i_k] \det\ A\,[\alpha; j_1, \ldots, j_{t-1}, i_s, j_{t+1}, \ldots, j_k] \end{aligned}$$

for each $s = 1, \ldots, k$ and all sequences of distinct indices $i_1, \ldots, i_k \in \{1, \ldots, n\}$ and $j_1, \ldots, j_k \in \{1, \ldots, n\}$.

0.8.9 The Laplace expansion theorem. The Laplace expansion (0.3.1.1) by minors along a given row or column is included in a natural family of expressions for the determinant. Let $A \in M_n(\mathbf{F})$, let $k \in \{1, \ldots, n\}$ be given, and let $\beta \subseteq \{1, \dots, n\}$ be any given index set of cardinality k. Then

$$\begin{aligned} \det A &= \sum_{\alpha} (-1)^{p(\alpha,\beta)} \det A[\alpha, \beta] \det A[\alpha^c, \beta^c] \\ &= \sum_{\alpha} (-1)^{p(\alpha,\beta)} \det A[\beta, \alpha] \det A[\beta^c, \alpha^c] \end{aligned}$$

in which the sums are over all index sets $\alpha \subseteq \{1, \dots, n\}$ of cardinality k, and $p(\alpha, \beta) \equiv \sum_{i \in \alpha} i + \sum_{j \in \beta} j$. Choosing $k = 1$ and $\beta = \{i\}$ or $\{j\}$ gives the expansions in (0.3.1.1).

0.8.10 Derivative of the determinant Let $A(t) = [a_1(t) \ \dots \ a_n(t)]$ be an n-by-n complex matrix whose entries are differentiable functions of t, partitioned according to its columns. It follows from multilinearity of the determinant (0.3.6(a)) and the definition of the derivative that

$$\frac{d}{dt} \det A(t) = \sum_{j=1}^{n} \det \left(A \underset{j}{\leftarrow} a_j'(t) \right) \tag{0.8.10.1}$$

For example, if $B = [b_1 \ \dots \ b_n] \in M_n$ is partitioned according to its columns and if $A(t) = tI - B = [te_1 - b_1 \ \dots \ te_n - b_n]$, then each $a_j'(t) = e_j$; (0.8.10.1) and a Laplace expansion by minors along column j identify the derivative of $\det(tI - A)$ as the sum of the principal minors of $tI - A$:

$$\frac{d}{dt} \det(tI - A) = \sum_{j=1}^{n} \det((tI - A)[\{j\}^c]) = \operatorname{tr} \operatorname{adj}(tI - A) \tag{0.8.10.2}$$

0.9 Special types of matrices

Certain matrices of special form arise frequently and have important properties. Some of these are cataloged here for reference and terminology.

0.9.1 Diagonal matrices. A matrix $D = [d_{ij}] \in M_{n,m}(\mathbf{F})$, is *diagonal* if $d_{ij} = 0$ whenever $j \neq i$. If all the diagonal entries of a diagonal matrix are positive (nonnegative) real numbers, we refer to it as a *positive* (*nonnegative*) *diagonal matrix*. The term *positive diagonal matrix* means that the matrix is diagonal and has positive diagonal entries; it does not refer to a general matrix with positive diagonal entries. The identity matrix $I \in M_n$ is a positive diagonal matrix. A square diagonal matrix D is a *scalar matrix* if its diagonal entries are all equal; that is, $D = \alpha I$ for some $\alpha \in \mathbf{F}$. Left or right multiplication of a matrix by a scalar matrix has the same effect as multiplying it by the corresponding scalar.

If $A = [a_{ij}] \in M_{n,m}(\mathbf{F})$ and $q = \min\{m, n\}$, then $\operatorname{diag} A = [a_{11}, \ldots, a_{qq}]^T \in \mathbf{F}^q$ denotes the vector of diagonal entries of A. Conversely, if $x \in \mathbf{F}^q$ and if m and n are positive integers such that $\min\{m, n\} = q$, then $\operatorname{diag} x \in M_{n,m}(\mathbf{F})$ denotes the n-by-m diagonal matrix A such that $\operatorname{diag} A = x$.

Suppose $D = [d_{ij}], E = [e_{ij}] \in M_n(\mathbf{F})$ are diagonal and let $A = [a_{ij}] \in M_n(\mathbf{F})$ be given. Then: (a) $\det D = \prod_{i=1}^n d_{ii}$; (b) D is nonsingular if and only if all $d_{ii} \neq 0$; (c) *Left* multiplication of A by D multiplies the *rows* of A by the diagonal entries of D (the ith row of DA is d_{ii} times the ith row of A); (d) *Right* multiplication of A by D multiplies the *columns* of A by the diagonal entries of D, that is, the jth column of AD is d_{jj} times the jth column of A; (e) $DA = AD$ if and only if $a_{ij} = 0$ whenever $d_{ii} \neq d_{jj}$; (f) If all the diagonal entries of D are distinct and $DA = AD$, then A is diagonal; (g) For any positive integer k, $D^k = \operatorname{diag}(d_{11}^k, \ldots, d_{nn}^k)$; (h) Any two diagonal matrices D and E of the same size commute: $DE = \operatorname{diag}(d_{11}e_{11}, \ldots, d_{nn}e_{nn}) = ED$.

0.9.2 Block diagonal matrices. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} A_{11} & & & 0 \\ & A_{22} & & \\ & & \ddots & \\ 0 & & & A_{kk} \end{bmatrix}$$

in which $A_{ii} \in M_{n_i}(\mathbf{F})$, $i = 1, \ldots, k$, and $\sum_{i=1}^k n_i = n$ is called *block diagonal.* It is convenient to write such a matrix as

$$A = A_{11} \oplus A_{22} \oplus \cdots \oplus A_{kk} \equiv \oplus_{i=1}^k A_{ii}$$

this is the *direct sum* of the matrices $A_{11}, \ldots, A_{kk}$. Many properties of block diagonal matrices generalize those of diagonal matrices. For example, det $(\oplus_{i=1}^k A_{ii}) = \prod_{i=1}^k \det A_{ii}$, so that $A = \oplus A_{ii}$ is nonsingular if and only if each A_{ii} is nonsingular, $i = 1, \ldots, k$. Furthermore, two direct sums $A = \oplus_{i=1}^k A_{ii}$ and $B = \oplus_{i=1}^k B_{ii}$, in which each A_{ii} is the same size as B_{ii}, commute if and only if each A_{ii} and B_{ii} commute, $i = 1, \ldots, k$. Also, $\operatorname{rank}(\oplus_{i=1}^k A_{ii}) = \sum_{i=1}^k \operatorname{rank} A_{ii}$.

0.9.3 Triangular matrices. A matrix $T = [t_{ij}] \in M_{n,m}(\mathbf{F})$ is *upper triangular* if $t_{ij} = 0$ whenever $i > j$. If $t_{ij} = 0$ whenever $i \geq j$, then T is said to be *strictly upper triangular*. Analogously, T is *lower triangular* (or *strictly lower triangular*) if its transpose is upper triangular (or strictly upper triangular). A *triangular* matrix is either lower or upper triangular. A square triangular matrix shares with a square diagonal matrix the property that its determinant is the product of its diagonal entries. Square triangular matrices need not commute with other square triangular matrices of the same size. However, if $T \in M_n$ is triangular, has distinct diagonal entries, and commutes with $B \in M_n$, then B must be triangular of the same type as T.

Left multiplication of $A \in M_n(\mathbf{F})$ by a lower triangular matrix L, that is, LA, replaces the ith row of A by a linear combination of the first through ith rows of A. Sometimes the terms *right* (in place of *upper*) and *left* (in place of *lower*) are used to describe triangular matrices. The rank of a triangular matrix is at least, and can be greater than, the number of nonzero entries on the main diagonal. If a square triangular matrix is nonsingular, its inverse is a triangular matrix of the same type. A product of square triangular matrices of the same size and type is a triangular matrix of the same type; each i, i diagonal entry of such a matrix product is the product of the i, i entries of the factors.

0.9.4 Block triangular matrices. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} A_{11} & & & \bigstar \\ & A_{22} & & \\ & & \ddots & \\ & 0 & & A_{kk} \end{bmatrix}$$

in which $A_{ii} \in M_{n_i}(\mathbf{F})$, $i = 1, \ldots, k$, $\Sigma_{i=1}^{k} n_i = n$, and all blocks below the block diagonal are zero, is *block upper triangular*; it is strictly block upper triangular if, in addition, all the diagonal blocks are zero blocks. A matrix is *block lower triangular* if its transpose is block upper triangular; it is *strictly block lower triangular* if its transpose is strictly block upper triangular. We say that a matrix is *block triangular* if it is either block lower triangular or block upper triangular; a matrix is both block lower triangular and block upper triangular if and only if it is block diagonal.

A block upper triangular matrix in which all the diagonal blocks are 1-by-1 or 2-by-2 is said to be *upper quasi-triangular*. A matrix is *lower quasi-triangular* if its transpose is upper quasi-triangular; it is *quasi-triangular* if it is either upper quasi-triangular or lower quasi-triangular. A matrix that is both upper quasi-triangular and lower quasi-triangular is said to be *quasi-diagonal*.

The determinant of a block triangular matrix is the product of the determinants of the diagonal blocks. The rank of a block triangular matrix is at least, and can be greater than, the sum of the ranks of the diagonal blocks. If a block triangular matrix is nonsingular, its inverse is a block triangular matrix of the same type with conformal blocks.

0.9.5 Permutation matrices. A square matrix P is a *permutation matrix* if exactly one entry in each row and column is equal to 1 and all other entries are 0. Multiplication by such matrices effects a permutation of the rows or columns of the matrix multiplied. For example,

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

illustrates how a permutation matrix produces a permutation of the rows (entries) of a vector: it sends the first entry to the second position, the second entry to the first position, and leaves the third entry in the third position. Left multiplication of a matrix $A \in M_{m,n}$ by an m-by-m permutation matrix P permutes the rows of A, while right multiplication of A by an n-by-n permutation matrix P permutes the columns of A. The matrix that carries out a Type 1 elementary operation (0.3.3) is an example of a special type of permutation matrix called a *transposition*. Any permutation matrix is a product of transpositions.

The determinant of a permutation matrix is ± 1, so permutation matrices are nonsingular. Although permutation matrices need not commute, the product of two permutation matrices is again a permutation matrix. Since the identity is a permutation matrix and $P^T = P^{-1}$ for every permutation matrix P, the permutation matrices constitute a subgroup of $GL(n, \mathbf{C})$ with cardinality $n!$.

Since right multiplication by $P^T = P^{-1}$ permutes columns in the same way that left multiplication by P permutes rows, the transformation $A \to PAP^T$ permutes the rows and columns (and the main diagonal entries) of $A \in M_n$ in the same way. In the context of linear equations with coefficient matrix A, this transformation amounts to renumbering the variables and the equations in the same way. A matrix $A \in M_n$ such that PAP^T is triangular for some permutation matrix P is called *essentially triangular;* these matrices have much in common with triangular matrices.

If $\Lambda \in M_n$ is diagonal and $P \in M_n$ is a permutation matrix, then $P\Lambda P^T$ is a diagonal matrix.

The n-by-n *reversal matrix* is the permutation matrix

$$K_n \equiv \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix} = [\kappa_{ij}] \in M_n \tag{0.9.5.1}$$

in which $\kappa_{n-i+1,i} = 1$ for $i = 1, \ldots, n$ and all other entries are zero. The rows of $K_n A$ are the rows of A presented in reverse order; the columns of AK_n are the columns of A presented in reverse order. The reversal matrix is sometimes called the *sip matrix* (*standard involutory permutation*), the *backwards identity*, or the *exchange matrix*.

For any n-by-n matrix $A = [a_{ij}]$, the entries $a_{n-i+1,i}$ for $i = 1, \ldots, n$ comprise its *backwards diagonal* (sometimes called the *secondary diagonal* or *anti-diagonal*).

0.9.6 Circulant matrices. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} a_1 & a_2 & \cdots\cdots\cdots & & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_2 & a_3 & \cdots & a_n & a_1 \end{bmatrix} \tag{0.9.6.1}$$

is a *circulant matrix*. Each row is the previous row cycled forward one step; the entries in each row are a cyclic permutation of those in the first. The n-by-n permutation matrix

$$C_n = \begin{bmatrix} 0 & 1 & 0 & .. & 0 \\ \vdots & 0 & 1 & & : \\ & & \ddots & \ddots & 0 \\ 0 & & & & 1 \\ 1 & 0 & & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 0 & I_{n-1} \\ 1 & 0_{1,n-1} \end{bmatrix} \tag{0.9.6.2}$$

is the *basic circulant permutation* matrix. A matrix $A \in M_n(\mathbf{F})$ can be written in the form

$$A = \sum_{k=0}^{n-1} a_{k+1} C_n^k \tag{0.9.6.3}$$

if and only if it is a circulant. We have $C_n^0 \equiv I = C_n^n$, and the coefficients $a_1, a_2, \ldots, a_n$ are the entries of the first row of A. Because of this representation, the circulant matrices of size n are a commutative algebra: linear combinations and products of circulants are circulants, and the inverse of a nonsingular circulant is a circulant; any two circulants of the same size commute.

0.9.7 Toeplitz matrices. A matrix $A = [a_{ij}] \in M_{n+1}(\mathbf{F})$ of the form

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \dots & \dots & a_n \\ a_{-1} & a_0 & a_1 & a_2 & \dots & a_{n-1} \\ a_{-2} & a_{-1} & a_0 & a_1 & \dots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & a_1 \\ a_{-n} & a_{-n+1} & \dots & \dots & a_{-1} & a_0 \end{bmatrix}$$

is a *Toeplitz matrix*. The entry a_{ij} is equal to a_{j-i} for some given sequence $a_{-n}, a_{-n+1}, \dots, a_{-1}, a_0, a_1, a_2, \dots, a_{n-1}, a_n \in \mathbf{C}$. The entries of A are constant down the diagonals parallel to the main diagonal. The Toeplitz matrices

$$B = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ & 0 & \ddots & \\ & & \ddots & 1 \\ \mathbf{0} & & & 0 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 0 & & & \mathbf{0} \\ 1 & 0 & & \\ & \ddots & \ddots & \\ \mathbf{0} & & 1 & 0 \end{bmatrix}$$

are called the *backward shift* and *forward shift* because of their effect on the elements of the standard basis $\{e_1, \dots, e_{n+1}\}$. A matrix $A \in M_{n+1}$ can be written in the form

$$A = \sum_{k=1}^{n} a_{-k} F^k + \sum_{k=0}^{n} a_k B^k$$

if and only if it is a Toeplitz matrix. Toeplitz matrices arise naturally in problems involving trigonometric moments.

An upper triangular Toeplitz matrix $A \in M_{n+1}(\mathbf{F})$ can be represented as a polynomial in B:

$$A = a_0 I + a_1 B + \cdots + a_n B^n$$

This representation (and the fact that $B^{n+1} = 0$) makes it clear why the upper triangular Toeplitz matrices of size n are a commutative algebra: linear combinations and products of upper triangular Toeplitz matrices are upper triangular Toeplitz matrices; A is nonsingular if and only if $a_0 \neq 0$, in which case $A^{-1} = b_0 I + b_1 B + \cdots + b_n B^n$ is also an upper triangular Toeplitz matrix with $b_0 = a_0^{-1}$ and $b_k = a_0^{-1}(\sum_{m=0}^{k-1} a_{k-m} b_m)$ for $k = 1, \dots, n$; any two upper triangular Toeplitz matrices of the same size commute.

0.9.8 Hankel matrices. A matrix $A \in M_{n+1}(\mathbf{F})$ of the form

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots\cdots & & a_n \\ a_1 & a_2 & a_3 & \cdots\cdots & a_n & a_{n+1} \\ a_2 & a_3 & \ddots & & & a_{n+2} \\ & \cdot & & \ddots & & \\ \vdots & & & & \ddots & \vdots \\ & a_n & & & & \\ a_n & a_{n+1} & a_{n+2} & \cdots\cdots & & a_{2n} \end{bmatrix}$$

is a *Hankel matrix*. Each entry a_{ij} is equal to a_{i+j-2} for some given sequence $a_0, a_1, a_2, \ldots, a_{2n-1}, a_{2n}$. The entries of A are constant along the diagonals perpendicular to the main diagonal. Hankel matrices arise naturally in problems involving power moments. Using the reversal matrix K of appropriate size (0.9.5.1), notice that KT and TK are Hankel matrices for any Toeplitz matrix T; KH and HK are Toeplitz matrices for any Hankel matrix H. Since $K = K^T = K^{-1}$ and Hankel matrices are symmetric, this means that any Toeplitz matrix is a product of two symmetric matrices with special structure: the reversal matrix and a Hankel matrix.

0.9.9 Hessenberg matrices. A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is said to be in *upper Hessenberg form* or to be an *upper Hessenberg matrix* if $a_{ij} = 0$ for all $i > j + 1$:

$$A = \begin{bmatrix} a_{11} & a_{12} & & \cdots & & a_{1n} \\ a_{21} & a_{22} & & & & \\ 0 & a_{32} & & \ddots & & \vdots \\ \vdots & 0 & \ddots & & & \vdots \\ & \vdots & \ddots & & & \\ 0 & 0 & \cdots & 0 & a_{n,n-1} & a_{nn} \end{bmatrix}$$

An upper Hessenberg matrix A is said to be *unreduced* if all its sub-diagonal entries are nonzero, that is, if $a_{i+1,i} \neq 0$ for all $i = 1, \ldots, n-1$; the rank of such a matrix is at least $n-1$ since its first $n-1$ columns are independent.

Let $A \in M_n(\mathbf{F})$ be unreduced upper Hessenberg. Then $A - \lambda I$ is unreduced upper Hessenberg for all $\lambda \in \mathbf{F}$, so $\operatorname{rank}(A - \lambda I) \geq n - 1$ for all $\lambda \in \mathbf{F}$.

A matrix $A \in M_n(\mathbf{F})$ is *lower Hessenberg* if A^T is upper Hessenberg.

0.9.10 Tridiagonal, bidiagonal, and other structured matrices. A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ that is *both* upper and lower Hessenberg is called *tridiagonal*, that is, A is tridiagonal if $a_{ij} = 0$ whenever $|i - j| > 1$:

$$A = \begin{bmatrix} a_1 & b_1 & & \\ c_1 & a_2 & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ & & c_{n-1} & a_n \end{bmatrix} \qquad (0.9.10.1)$$

The determinant of A can be calculated inductively starting with $\det A_1 = a_1$, $\det A_2 = a_1a_2 - b_1c_1$, and then computing a sequence of 2-by-2 matrix products

$$\begin{bmatrix} \det A_{k+1} & 0 \\ \det A_k & 0 \end{bmatrix} = \begin{bmatrix} a_{k+1} & -b_kc_k \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \det A_k & 0 \\ \det A_{k-1} & 0 \end{bmatrix}, \quad k = 2, \dots, n-1$$

An *upper bidiagonal matrix* $A \in M_n(\mathbf{F})$ is a tridiagonal matrix (0.9.10.1) in which $c_1 = \cdots = c_{n-1} = 0$. A matrix $A \in M_n(\mathbf{F})$ is *lower bidiagonal* if A^T is upper bidiagonal.

A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is *persymmetric* if $a_{ij} = a_{n+1-j,n+1-i}$ for all $i, j = 1, \dots, n$, that is, A is persymmetric if $A = K_nA^TK_n$, in which K_n is the reversal matrix (0.9.5.1); A is *skew-persymmetric* if $A = -K_nA^TK_n$. Thus, a persymmetric matrix is symmetric with respect to the backwards diagonal. If A is persymmetric and invertible, then A^{-1} is also persymmetric since $A^{-1} = K_n(A^{-1})^TK_n$. Toeplitz matrices are persymmetric. A complex matrix A such that $A = K_nA^*K_n$ is said to be *perhermitian*; the inverse of a perhermitian and invertible matrix is perhermitian.

A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is called *centrosymmetric* if $a_{ij} = a_{n+1-i,n+1-j}$ for all $i, j = 1, \dots, n$, that is, A is centrosymmetric if $A = K_nAK_n$; A is *skew-centrosymmetric* if $A = -K_nAK_n$. A centrosymmetric matrix is symmetric about its geometric center, as illustrated by the example

$$A = \begin{bmatrix} 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & 3 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 1 & 0 \\ 1 & 1 & 2 & 0 & 1 \end{bmatrix}$$

If A is centrosymmetric and invertible, then A^{-1} is also centrosymmetric, since $A^{-1} = K_nA^{-1}K_n$. A complex matrix A such that $A = K_n\bar{A}K_n$ is said to be *centrohermitian*; the inverse of a centrohermitian and invertible matrix is centrohermitian.

0.9.11 Vandermonde matrices and Lagrange interpolation. A *Vandermonde matrix* $A \in M_n(\mathbf{F})$ is a matrix of the form

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix} \tag{0.9.11.1}$$

in which $x_1, x_2, \dots, x_n \in \mathbf{F}$; that is, $A = [a_{ij}]$ with $a_{ij} = x_i^{j-1}$. It is a fact that

$$\det A = \prod_{\substack{i,j=1 \\ i>j}}^{n} (x_i - x_j) \tag{0.9.11.2}$$

so a Vandermonde matrix is nonsingular if and only if the *n* parameters $x_1, x_2, \dots, x_n$ are distinct.

The Vandermonde matrix arises in the *interpolation problem* of finding a polynomial $p(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1 x + a_0$ of degree at most $n-1$ with coefficients from **F** such that

$$\begin{aligned} p(x_1) &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} = y_1 \\ p(x_2) &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} = y_2 \\ &\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ p(x_n) &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_{n-1} x_n^{n-1} = y_n \end{aligned} \tag{0.9.11.3}$$

in which $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$ are given elements of **F**. The interpolation conditions (0.9.11.3) are a system of *n* equations for the *n* unknown coefficients $a_0, a_1, \dots, a_{n-1}$, and they have the form $Aa = y$, in which $a = [a_0, a_1, \dots, a_{n-1}]^T \in \mathbf{F}^n$, $y = [y_1, y_2, \dots, y_n]^T \in \mathbf{F}^n$, and $A \in M_n(\mathbf{F})$ is the Vandermonde matrix (0.9.11.1). This interpolation problem always has a solution if the points $x_1, x_2, \dots, x_n$ are distinct, for *A* is nonsingular in this event.

If the points $x_1, x_2, \dots, x_n$ are distinct, the coefficients of the interpolating polynomial could in principle be obtained by solving the system (0.9.11.3), but it is usually more useful to represent the interpolating polynomial *p(x)* as a linear combination of the *Lagrange interpolating polynomials*

$$L_i(x) = \frac{\prod_{\substack{j=1 \\ j\neq 1}}^{n} (x - x_j)}{\prod_{\substack{j=1 \\ j\neq 1}}^{n} (x_i - x_j)}, \qquad i = 1, 2, \dots, n$$

Each polynomial $L_i(x)$ has degree $n-1$ and has the property that $L_i(x_k)=0$ if $k \neq i$, but $L_i(x_i)=1$. *Lagrange's interpolation formula*

$$p(x) = y_1 L_1(x) + y_2 L_2(x) + \cdots + y_n L_n(x) \tag{0.9.11.4}$$

provides a polynomial of degree at most $n-1$ that satisfies the equations (0.9.11.3).

0.9.12 Cauchy matrices A *Cauchy matrix* $A \in M_n(\mathbf{F})$ is a matrix of the form $A = [(a_i + b_i)^{-1}]_{i,j=1}^n$, in which $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ are scalars such that $a_i + b_j \neq 0$ for all $i, j = 1, \ldots, n$. It is a fact that

$$\det A = \frac{\prod\limits_{1 \le i < j \le n} (a_j - a_i)(b_j - b_i)}{\prod\limits_{1 \le i \le j \le n} (a_i + b_j)} \tag{0.9.12.1}$$

so A is nonsingular if and only if $a_i \neq a_j$ and $b_i \neq b_j$ for all $i \neq j$. A *Hilbert matrix* $H = [(i+j-1)^{-1}]_{i,j=1}^n$ is a Cauchy matrix that is also a Hankel matrix. It is a fact that

$$\det H = \frac{(1!2!\cdots(n-1)!)^4}{1!2!\cdots(2n-1)!} \tag{0.9.12.2}$$

so a Hilbert matrix is always nonsingular.

0.9.13 Involution, nilpotent, projection, coninvolution A matrix $A \in M_n(\mathbf{F})$ is

- an *involution* if $A^2 = I$, that is, if $A = A^{-1}$.
- *nilpotent* if $A^k = 0$ for some positive integer k; the least such k is the *index of nilpotence* of A.
- a *projection* if $A^2 = A$ (the term *idempotent* is also used).

Now suppose that $\mathbf{F} = \mathbf{C}$. A matrix $A \in M_n$ is

- a *Hermitian projection* if $A^* = A$ and $A^2 = A$ (the term *orthogonal projection* is also used).
- *coninvolutory* if $A\bar{A} = I$, that is, if $\bar{A} = A^{-1}$.

0.10 Change of basis

Let V be an *n*-dimensional vector space over the field **F**, and let $\mathcal{B}_1 = \{v_1, v_2, \ldots, v_n\}$ be a basis for *V*. Any vector $x \in V$ can be represented as $x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ because $\mathcal{B}_1$ spans *V*. If there were some other representation of $x = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n$ in the same basis, then

$$0 = x - x = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n$$

from which it follows that all $\alpha_i - \beta_i = 0$ because the set $\mathcal{B}_1$ is independent. Given the basis $\mathcal{B}_1$, the linear mapping

$$x \to [x]_{\mathcal{B}_1} \equiv \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}, \qquad \text{in which } x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

from V to $\mathbf{F}^n$ is well defined, one-to-one, and onto. The scalars α_i are the *coordinates* of *x* with respect to the basis $\mathcal{B}_1$, and the column vector $[x]_{\mathcal{B}_1}$ is the unique $\mathcal{B}_1$*-coordinate representation of x*.

Let $T : V \to V$ be a given linear transformation. The action of T on any $x \in V$ is determined once one knows the *n* vectors $Tv_1, Tv_2, \ldots, Tv_n$, because any $x \in V$ has a unique representation $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $Tx = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = T(\alpha_1 v_1) + \cdots + T(\alpha_n v_n) = \alpha_1 T v_1 + \cdots + \alpha_n T v_n$ by linearity. Thus, the value of *Tx* is determined once $[x]_{\mathcal{B}_1}$ is known.

Let $\mathcal{B}_2 = \{w_1, w_2, \ldots, w_n\}$ also be a basis for V (either different from or the same as $\mathcal{B}_1$) and suppose that the $\mathcal{B}_2$-coordinate representation of Tv_j is

$$[Tv_j]_{\mathcal{B}_2} = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{nj} \end{bmatrix}, \qquad j = 1, 2, \ldots, n$$

Then for any $x \in V$ we have

$$\begin{aligned} [Tx]_{\mathcal{B}_2} &= \left[\sum_{j=1}^{n} \alpha_j T v_j \right]_{\mathcal{B}_2} = \sum_{j=1}^{n} \alpha_j \, [Tv_j]_{\mathcal{B}_2} \\ &= \sum_{j=1}^{n} \alpha_j \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{nj} \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \end{aligned}$$

The *n*-by-*n* array $[t_{ij}]$ depends on T and on the choice of the bases $\mathcal{B}_1$ and $\mathcal{B}_2$, but it does not depend on *x*. We define the $\mathcal{B}_1$-$\mathcal{B}_2$ *basis representation of T* to be

$$_{\mathcal{B}_2}[T]_{\mathcal{B}_1} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} = \left[[Tv_1]_{\mathcal{B}_2} \ldots [Tv_n]_{\mathcal{B}_2}\right]$$

We have just shown that $[Tx]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[T]_{\mathcal{B}_1}[x]_{\mathcal{B}_1}$ for any $x \in V$. In the important special case $\mathcal{B}_2 = \mathcal{B}_1$, we have $_{\mathcal{B}_1}[T]_{\mathcal{B}_1}$, which is called the $\mathcal{B}_1$ *basis representation of T*.

Consider the identity linear transformation $I : V \to V$ defined by $Ix = x$ for all *x*. Then

$$[x]_{\mathcal{B}_2} = [Ix]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[Ix]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}[x]_{\mathcal{B}_2}$$

for all $x \in V$. By successively choosing $x = w_1, w_2, \ldots, w_n$, this identity permits us to identify each column of $_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ and shows that

$$_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} = I_n$$

If we do the same calculation starting with $[x]_{\mathcal{B}_1} = [I]x_{\mathcal{B}_1} = \cdots$, we find that

$$_{\mathcal{B}_1}[I]_{\mathcal{B}_2}\,{}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = I_n$$

Thus, every matrix of the form $_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is invertible and $_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ is its inverse. Conversely, every invertible matrix $S = [s_1\ s_2\ \ldots\ s_n] \in M_n(\mathbf{F})$ has the form $_{\mathcal{B}_1}[I]_{\mathcal{B}}$ for some basis $\mathcal{B}$. We may take $\mathcal{B}$ to be the vectors $\{\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n\}$ defined by $[\tilde{s}_i]_{\mathcal{B}_1} = s_i$, $i = 1, 2, \ldots, n$. The set $\mathcal{B}$ is independent because *S* is invertible.

Notice that

$$_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = \left[[Iv_1]_{\mathcal{B}_2} \ldots [Iv_n]_{\mathcal{B}_2}\right] = \left[[v_1]_{\mathcal{B}_2} \ldots [v_n]_{\mathcal{B}_2}\right]$$

so $_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ expresses the elements of the basis $\mathcal{B}_1$ in terms of the basis $\mathcal{B}_2$. Now let $x \in V$ and compute

$$\begin{aligned} {}_{\mathcal{B}_1}[T]_{\mathcal{B}_2}[x]_{\mathcal{B}_2} = [Tx]_{\mathcal{B}_2} &= [I(Tx)]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[Tx]_{\mathcal{B}_1} \\ &= {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}[x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}[Ix]_{\mathcal{B}_1} \\ &= {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}[x]_{\mathcal{B}_2} \end{aligned}$$

By choosing $x = w_1, w_2, \ldots, w_n$ successively, we conclude that

$$_{\mathcal{B}_2}[T]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} \tag{0.10.1.1}$$

This identity shows how the $\mathcal{B}_1$ basis representation of *T* changes if the basis is changed to $\mathcal{B}_2$. For this reason, the matrix ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is called the $\mathcal{B}_1 - \mathcal{B}_2$ *change of basis matrix*.

Any matrix $A \in M_n(\mathbf{F})$ is a basis representation of some linear transformation $T : V \to V$, for if $\mathcal{B}$ is any basis of V, we can determine *Tx* by $[Tx]_{\mathcal{B}} = A[x]_{\mathcal{B}}$. For this *T*, a computation reveals that ${}_{\mathcal{B}}[T]_{\mathcal{B}} = A$.

0.11 Equivalence Relations

Let S be a given set and let Γ be a given subset of $S \times S \equiv \{(a, b) : a \in S$ and $b \in S\}$. Then Γ defines a *relation* on S in the following way: we say that a is *related to* b, written $a \sim b$, if $(a, b) \in \Gamma$. A relation on S is said to be an *equivalence relation* if it is: (a) *reflexive* ($a \sim a$ for every $a \in S$), (b) *symmetric* ($a \sim b$ whenever $b \sim a$), and (c) *transitive* ($a \sim c$ whenever $a \sim b$ and $b \sim c$). An equivalence relation on S gives a disjoint partition of S in a natural way: If we define the *equivalence class* of any $a \in S$ by $S_a \equiv \{b \in S : b \sim a\}$, then $S = \cup_{a \in S} S_a$ and for each $a, b \in S$, either $S_a = S_b$ (if $a \sim b$) or $S_a \cap S_b = \emptyset$ (if $a \not\sim b$). Conversely, any disjoint partition of S can be used to define an equivalence relation on S.

The following table lists several equivalence relations that play important roles in matrix analysis. The factors S and T are nonsingular, and the factors U and V are unitary; A and B need not be square for equivalence or unitary equivalence.

Equivalence Relation $\sim$	$A \sim B$
congruence	$A = SBS^T$
unitary congruence	$A = UBU^T$
*congruence	$A = SBS^*$
consimilarity	$A = SB\bar{S}^{-1}$
equivalence	$A = SBT$
unitary equivalence	$A = UBV$
similarity	$A = SBS^{-1}$
unitary similarity	$A = UBU^*$

Whenever an interesting equivalence relation arises in matrix analysis, it can be useful to identify a distinguished representative of each equivalence class (a *canonical form* or *normal form*). Alternatively, we often want to have effective criteria (*invariants*) that can be used to decide if two given matrices belong to the same equivalence class.

Abstractly, a *canonical form* for an equivalence relation $\sim$ on S is a subset C of S such that $S = \cup_{a \in C} S_a$ and $S_a \cap S_b = \emptyset$ whenever $a, b \in C$ and $a \neq b$. For a given equivalence relation in matrix analysis, it is important to make an artful and simple choice of canonical form, and one sometimes does this in more than one way to tailor the canonical form to a specific purpose. For example, the Jordan and Weyr canonical forms are different canonical forms for similarity; the Jordan canonical form works well in problems involving powers of matrices, while the Weyr canonical form works well in problems involving commutativity.

An *invariant* for an equivalence relation $\sim$ on S is a function f on S such that $f(a) = f(b)$ whenever $a \sim b$. A family of invariants $\mathcal{F}$ for an equivalence relation $\sim$ on S is said to be *complete* if $f(a) = f(b)$ for all $f \in \mathcal{F}$ if and only if $a \sim b$; a complete family of invariants is often called a *complete system of invariants*. For example, the singular values of a matrix are a complete system of invariants for unitary equivalence.

Index