II. Vector Spaces over $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, 33-88

DOI: 10.3792/euclid/9781429799980-2

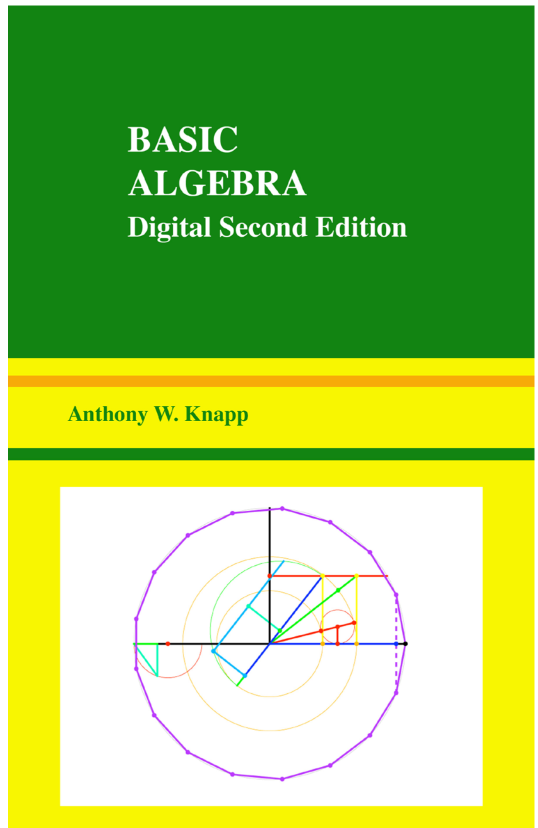
from

Basic Algebra

Digital Second Edition

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799980
ISBN: 978-1-4297-9998-0



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/∼aknapp

Title: Basic Algebra
Cover: Construction of a regular heptadecagon, the steps shown in color sequence; see page 505.

Mathematics Subject Classification (2010): 15–01, 20–01, 13–01, 12–01, 16–01, 08–01, 18A05, 68P30.

First Edition, ISBN-13 978-0-8176-3248-9

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the Author

CHAPTER II

Vector Spaces over $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$

Abstract. This chapter introduces vector spaces and linear maps between them, and it goes on to develop certain constructions of new vector spaces out of old, as well as various properties of determinants.

Sections 1–2 define vector spaces, spanning, linear independence, bases, and dimension. The sections make use of row reduction to establish dimension formulas for certain vector spaces associated with matrices. They conclude by stressing methods of calculation that have quietly been developed in proofs.

Section 3 relates matrices and linear maps to each other, first in the case that the linear map carries column vectors to column vectors and then in the general finite-dimensional case. Techniques are developed for working with the matrix of a linear map relative to specified bases and for changing bases. The section concludes with a discussion of isomorphisms of vector spaces.

Sections 4–6 take up constructions of new vector spaces out of old ones, together with corresponding constructions for linear maps. The four constructions of vector spaces in these sections are those of the dual of a vector space, the quotient of two vector spaces, and the direct sum and direct product of two or more vector spaces.

Section 7 introduces determinants of square matrices, together with their calculation and properties. Some of the results that are established are expansion in cofactors, Cramer's rule, and the value of the determinant of a Vandermonde matrix. It is shown that the determinant function is well defined on any linear map from a finite-dimensional vector space to itself.

Section 8 introduces eigenvectors and eigenvalues for matrices, along with their computation. Also, in this section the characteristic polynomial and the trace of a square matrix are defined, and all these notions are reinterpreted in terms of linear maps.

Section 9 proves the existence of bases for infinite-dimensional vector spaces and discusses the extent to which the material of the first eight sections extends from the finite-dimensional case to be valid in the infinite-dimensional case.

1. Spanning, Linear Independence, and Bases

This chapter develops a theory of rational, real, and complex vector spaces. Many readers will already be familiar with some aspects of this theory, particularly in the case of the vector spaces $\mathbb{Q}^n$, $\mathbb{R}^n$, and $\mathbb{C}^n$ of column vectors, where the tools developed from row reduction allow one to introduce geometric notions and to view geometrically the set of solutions to a set of linear equations. Thus we shall be brief about many of these matters, concentrating on the algebraic aspects of the theory. Let $\mathbb{F}$ denote any of $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$. Members of $\mathbb{F}$ are called **scalars**.[1]

A **vector space** over $\mathbb{F}$ is a set V with two operations, **addition** carrying $V \times V$ into V and **scalar multiplication** carrying $\mathbb{F} \times V$ into V, with the following properties:

(i) the operation of addition, written $+$, satisfies

(a) $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$ for all v_1, v_2, v_3 in V (associative law),
(b) there exists an element 0 in V with $v + 0 = 0 + v = v$ for all v in V,
(c) to each v in V corresponds an element $-v$ in V such that $v + (-v) = (-v) + v = 0$,
(d) $v_1 + v_2 = v_2 + v_1$ for all v_1 and v_2 in V (commutative law);

(ii) the operation of scalar multiplication, written without a sign, satisfies

(a) $a(bv) = (ab)v$ for all v in V and all scalars a and b,
(b) $1v = v$ for all v in V and for the scalar 1;

(iii) the two operations are related by the distributive laws

(a) $a(v_1 + v_2) = av_1 + av_2$ for all v_1 and v_2 in V and for all scalars a,
(b) $(a + b)v = av + bv$ for all v in V and all scalars a and b.

It is immediate from these properties that

- 0 is unique (since $0' = 0' + 0 = 0$),
- $-v$ is unique (since $(-v)' = (-v)' + 0 = (-v)' + (v + (-v)) = ((-v)' + v) + (-v) = 0 + (-v) = (-v)$),
- $0v = 0$ (since $0v = (0 + 0)v = 0v + 0v$),
- $(-1)v = -v$ (since $0 = 0v = (1+(-1))v = 1v+(-1)v = v+(-1)v$),
- $a0 = 0$ (since $a0 = a(0 + 0) = a0 + a0$).

Members of V are called **vectors**.

EXAMPLES.

(1) $V = M_{kn}(\mathbb{F})$, the space of all k-by-n matrices. The above properties of a vector space over $\mathbb{F}$ were already observed in Section I.6. The vector space $\mathbb{F}^k$ of all k-dimensional column vectors is the special case $n = 1$, and the vector space $\mathbb{F}$ of scalars is the special case $k = n = 1$.

(2) Let S be any nonempty set, and let V be the set of all functions from S into $\mathbb{F}$. Define operations by $(f + g)(s) = f(s) + g(s)$ and $(cf)(s) = c(f(s))$. The operations on the right sides of these equations are those in $\mathbb{F}$, and the properties of a vector space follow from the fact that they hold in $\mathbb{F}$ at each s.

[1] All the material of this chapter will ultimately be seen to work when $\mathbb{F}$ is replaced by any "field." This point will not be important for us at this stage, and we postpone considering it further until Chapter IV.

(3) More generally than in Example 2, let S be any nonempty set, let U be a vector space over $\mathbb{F}$, and let V be the set of all functions from S into U. Define the operations as in Example 2, but interpret the operations on the right sides of the defining equations as those in U. Then the properties of a vector space follow from the fact that they hold in U at each s.

(4) Let V be any vector space over $\mathbb{C}$, and restrict scalar multiplication to an operation $\mathbb{R} \times V \to V$. Then V becomes a vector space over $\mathbb{R}$. In particular, $\mathbb{C}$ is a vector space over $\mathbb{R}$.

(5) Let $V = \mathbb{F}[X]$ be the set of all polynomials in one indeterminate with coefficients in $\mathbb{F}$, and define addition and scalar multiplication as in Section I.3. Then V is a vector space.

(6) Let V be any vector space over $\mathbb{F}$, and let U be any nonempty subset closed under addition and scalar multiplication. Then U is a vector space over $\mathbb{F}$. Such a subset U is called a **vector subspace** of V; sometimes one says simply **subspace** if the context is unambiguous.[2]

(7) Let V be any vector space over $\mathbb{F}$, and let $U = \{v_\alpha\}$ be any subset of V. A **finite linear combination** of the members of U is any vector of the form $c_{\alpha_1} v_{\alpha_1} + \cdots + c_{\alpha_n} v_{\alpha_n}$ with each c_{α_j} in $\mathbb{F}$, each v_{α_j} in U, and $n \geq 0$. The **linear span** of U is the set of all finite linear combinations of members of U. It is a vector subspace of V and is denoted by $\operatorname{span}\{v_\alpha\}$. By convention, $\operatorname{span} \varnothing = 0$.

(8) Many vector subspaces arise in the context of some branch of mathematics after some additional structure is imposed. For example let V be the vector space of all functions from $\mathbb{R}^3$ into $\mathbb{R}$, an instance of Example 2. The subset U of continuous members of V is a vector subspace; the closure under addition and scalar multiplication comes down to knowing that addition is a continuous function from $\mathbb{R}^3 \times \mathbb{R}^3$ into $\mathbb{R}^3$ and that scalar multiplication from $\mathbb{R} \times \mathbb{R}^3$ into $\mathbb{R}^3$ is continuous as well. Another example is the subset of twice continuously differentiable members f of V satisfying the partial differential equation $\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2} + f = 0$ on $\mathbb{R}^3$.

The associative and commutative laws in the definition of "vector space" imply certain more complicated formulas of which the stated laws are special cases. With associativity of addition, if n vectors $v_1, \ldots, v_n$ are given, then any way of inserting parentheses into the expression $v_1 + v_2 + \cdots + v_n$ leads to the same result, and a similar conclusion applies to the associativity-like formula $a(bv) = (ab)v$ for scalar multiplication. In the presence of associativity, the commutative law for addition implies that $v_1 + v_2 + \cdots + v_n = v_{\sigma(1)} + v_{\sigma(2)} + \cdots + v_{\sigma(n)}$ for any

[2] The word "subspace" arises also in the context of metric spaces and more general topological spaces, and the metric-topological notion of subspace is distinct from the vector notion of subspace.

permutation of $\{1, \ldots, n\}$. All these facts are proved by inductive arguments, and the details are addressed in Problems 2–3 at the end of the chapter.

Let V be a vector space over $\mathbb{F}$. A subset $\{v_\alpha\}$ of V **spans** V or is a **spanning set** for V if the linear span of $\{v_\alpha\}$, in the sense of Example 7 above, is all of V. A subset $\{v_\alpha\}$ is **linearly independent** if whenever a finite linear combination $c_{\alpha_1} v_{\alpha_1} + \cdots + c_{\alpha_n} v_{\alpha_n}$ equals the 0 vector, then all the coefficients must be 0: $c_{\alpha_1} = \cdots = c_{\alpha_n} = 0$. By subtraction we see that in this case any equality of two finite linear combinations

$$c_{\alpha_1} v_{\alpha_1} + \cdots + c_{\alpha_n} v_{\alpha_n} = d_{\alpha_1} v_{\alpha_1} + \cdots + d_{\alpha_n} v_{\alpha_n}$$

implies that the respective coefficients are equal: $c_{\alpha_j} = d_{\alpha_j}$ for $1 \leq j \leq n$.

A subset $\{v_\alpha\}$ is a **basis** if it spans V and is linearly independent. In this case each member of V has one and only one expansion as a finite linear combination of the members of $\{v_\alpha\}$.

EXAMPLE. In $\mathbb{F}^n$, the vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \qquad \ldots, \qquad e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

form a basis of $\mathbb{F}^n$ called the **standard basis** of $\mathbb{F}^n$.

Proposition 2.1. Let V be a vector space over $\mathbb{F}$.

(a) If $\{v_\alpha\}$ is a linearly independent subset of V that is maximal with respect to the property of being linearly independent (i.e., has the property of being strictly contained in no linearly independent set), then $\{v_\alpha\}$ is a basis of V.

(b) If $\{v_\alpha\}$ is a spanning set for V that is minimal with respect to the property of spanning (i.e., has the property of strictly containing no spanning set), then $\{v_\alpha\}$ is a basis of V.

PROOF. For (a), let v be given. We are to show that v is in the span of $\{v_\alpha\}$. Without loss of generality, we may assume that v is not in the set $\{v_\alpha\}$ itself. By the assumed maximality, $\{v_\alpha\} \cup \{v\}$ is not linearly independent, and hence $cv + c_{\alpha_1} v_{\alpha_1} + \cdots + c_{\alpha_n} v_{\alpha_n} = 0$ for some scalars $c, c_{\alpha_1}, \ldots, c_{\alpha_n}$ not all 0. Here $c \neq 0$ since $\{v_\alpha\}$ is linearly independent. Then $v = -c^{-1} c_{\alpha_1} v_{\alpha_1} - \cdots - c^{-1} c_{\alpha_n} v_{\alpha_n}$, and v is exhibited as in the linear span of $\{v_\alpha\}$.

For (b), suppose that $c_{\alpha_1} v_{\alpha_1} + \cdots + c_{\alpha_n} v_{\alpha_n} = 0$ with $c_{\alpha_1}, \ldots, c_{\alpha_n}$ not all 0. Say $c_{\alpha_1} \neq 0$. Then we can solve for v_{α_1} and see that v_{α_1} is a finite linear combination of $v_{\alpha_2}, \ldots, v_{\alpha_n}$. Substitution shows that any finite linear combination of the v_α's is a finite linear combination of the v_α's other than v_{α_1}, and we obtain a contradiction to the assumed minimality of the spanning set. $\square$

Proposition 2.2. Let V be a vector space over $\mathbb{F}$. If V has a finite spanning set $\{v_1, \dots, v_m\}$, then any linearly independent set in V has $\leq m$ elements.

PROOF. It is enough to show that no subset of $m+1$ vectors can be linearly independent. Arguing by contradiction, suppose that $\{u_1, \dots, u_n\}$ is a linearly independent set with $n = m+1$. Write

$$\begin{aligned} u_1 &= c_{11}v_1 + c_{21}v_2 + \cdots + c_{m1}v_m, \\ &\vdots \\ u_n &= c_{1n}v_1 + c_{2n}v_2 + \cdots + c_{mn}v_m. \end{aligned}$$

The system of linear equations

$$\begin{aligned} c_{11}x_1 + \cdots + c_{1n}x_n &= 0, \\ &\vdots \\ c_{m1}x_1 + \cdots + c_{mn}x_n &= 0, \end{aligned}$$

is a homogeneous system of linear equations with more unknowns than equations, and Proposition 1.26d shows that it has a nonzero solution $(x_1, \dots, x_n)$. Then we have

$$\begin{aligned} x_1u_1 + \cdots + x_nu_n = {} & c_{11}x_1v_1 + c_{21}x_1v_2 + \cdots + c_{m1}x_1v_m \\ & \quad + \qquad\quad + \qquad\qquad\quad + \\ & \quad \cdots \qquad\quad \cdots \qquad\qquad\quad \cdots \\ & \quad + \qquad\quad + \qquad\qquad\quad + \\ & c_{1n}x_nv_1 + c_{2n}x_nv_2 + \cdots + c_{mn}x_nv_m \\ = {} & 0, \end{aligned}$$

in contradiction to the assumed linear independence of $\{u_1, \dots, u_n\}$. $\square$

Corollary 2.3. If the vector space V has a finite spanning set $\{v_1, \dots, v_m\}$, then

(a) $\{v_1, \dots, v_m\}$ has a subset that is a basis,
(b) any linearly independent set in V can be extended to a basis,
(c) V has a basis,
(d) any two bases have the same finite number of elements, necessarily $\leq m$.

REMARKS. In this case we say that V is **finite-dimensional**, and the number of elements in a basis is called the **dimension** of V, written $\dim V$. If V has no finite spanning set, we say that V is infinite-dimensional. A suitable analog of the conclusion in Corollary 2.3 is valid in the infinite-dimensional case, but the proof is more complicated. We take up the infinite-dimensional case in Section 9.

PROOF. By discarding elements of the set $\{v_1, \ldots, v_m\}$ one at a time if necessary and by applying Proposition 2.1b, we obtain (a). For (b), we see from Proposition 2.2 that the given linearly independent set has $\leq m$ elements. If we adjoin elements to it one at a time so as to obtain larger linearly independent sets, Proposition 2.2 shows that there must be a stage at which we can proceed no further without violating linear independence. Proposition 2.1a then says that we have a basis. For (c), we observe that (a) has already produced a basis. Any two bases have the same number of elements, by two applications of Proposition 2.2, and this proves (d). $\square$

EXAMPLES. The vector space $M_{kn}(\mathbb{F})$ of k-by-n matrices has dimension kn. The vector space of all polynomials in one indeterminate is infinite-dimensional because the subspace consisting of 0 and of all polynomials of degree $\leq n$ has dimension $n + 1$.

Corollary 2.4. If V is a finite-dimensional vector space with $\dim V = n$, then any spanning set of n elements is a basis of V, and any linearly independent set of n elements is a basis of V. Consequently any n-dimensional vector subspace U of V coincides with V.

PROOF. These conclusions are immediate from parts (a) and (b) of Corollary 2.3 if we take part (d) into account. $\square$

Corollary 2.5. If V is a finite-dimensional vector space and U is a vector subspace of V, then U is finite-dimensional, and $\dim U \leq \dim V$.

PROOF. Let $\{v_1, \ldots, v_m\}$ be a basis of V. According to Proposition 2.2, any linearly independent set in U has $\leq m$ elements, being linearly independent in V. We can thus choose a maximal linearly independent subset of U with $\leq m$ elements, and Proposition 2.1a shows that the result is a basis of U. $\square$

2. Vector Spaces Defined by Matrices

Let A be a member of $M_{kn}(\mathbb{F})$, thus a k-by-n matrix. The **row space** of A is the linear span of the rows of A, regarded as a vector subspace of the vector space of all n-dimensional row vectors. The **column space** of A is the linear span of the columns, regarded as a vector subspace of k-dimensional column vectors. The **null space** of A is the vector subspace of n-dimensional column vectors v for which $Av = 0$, where Av is the matrix product. The fact that this last space is a vector subspace follows from the properties $A(v_1 + v_2) = Av_1 + Av_2$ and $A(cv) = c(Av)$ of matrix multiplication.

We can use matrix multiplication to view the matrix A as defining a function $v \mapsto Av$ of $\mathbb{F}^n$ to $\mathbb{F}^k$. This function satisfies the properties just listed,

$$A(v_1 + v_2) = Av_1 + Av_2 \qquad \text{and} \qquad A(cv) = c(Av),$$

and we shall consider further functions with these two properties starting in the next section. In terms of this function, the null space of A is the set in the domain $\mathbb{F}^n$ mapped to 0. Because of these same properties and because the product Ae_j of A and the j^{th} standard basis vector e_j in $\mathbb{F}^n$ is the j^{th} column of A, the column space of A is the image of the function $v \mapsto Av$ as a subset of the range $\mathbb{F}^k$.

Theorem 2.6. If A is in $M_{kn}(\mathbb{F})$, then

$$\dim(\text{column space}(A)) + \dim(\text{null space}(A)) = \#(\text{columns of } A) = n.$$

PROOF. Corollary 2.5 says that the null space is finite-dimensional, being a vector subspace of $\mathbb{F}^n$, and Corollary 2.3c shows that the null space has a basis, say $\{v_1, \dots, v_r\}$. By Corollary 2.3b we can adjoin vectors $v_{r+1}, \dots, v_n$ so that $\{v_1, \dots, v_n\}$ is a basis of $\mathbb{F}^n$. If v is in $\mathbb{F}^n$, we can expand v in terms of this basis as $v = c_1v_1 + \cdots + c_nv_n$. Application of A gives

$$\begin{aligned} Av = A(c_1v_1 + \cdots + c_nv_n) &= c_1Av_1 + \cdots + c_rAv_r + c_{r+1}Av_{r+1} + \cdots + c_nAv_n \\ &= c_{r+1}Av_{r+1} + \cdots + c_nAv_n. \end{aligned}$$

Therefore the vectors $Av_{r+1}, \dots, Av_n$ span the column space.

Let us see that they form a basis for the column space. Thus suppose that $c_{r+1}Av_{r+1} + \cdots + c_nAv_n = 0$. Then $A(c_{r+1}v_{r+1} + \cdots + c_nv_n) = 0$, and $c_{r+1}v_{r+1} + \cdots + c_nv_n$ is in the null space. Since $\{v_1, \dots, v_r\}$ is a basis of the null space, we have

$$c_{r+1}v_{r+1} + \cdots + c_nv_n = a_1v_1 + \cdots + a_rv_r$$

for suitable scalars $a_1, \dots, a_r$. Therefore

$$(-a_1)v_1 + \cdots + (-a_r)v_r + c_{r+1}v_{r+1} + \cdots + c_nv_n = 0.$$

Since $v_1, \dots, v_n$ are linearly independent, all the c_j are 0. We conclude that $Av_{r+1}, \dots, Av_n$ are linearly independent and therefore form a basis of the column space.

As a result, we have established in the identity $r + (n - r) = n$ that $n - r$ can be interpreted as $\dim(\text{column space}(A))$ and that r can be interpreted as $\dim(\text{null space}(A))$. The theorem follows. $\square$

Proposition 2.7. If A is in $M_{kn}(\mathbb{F})$, then each elementary row operation on A preserves the row space of A.

PROOF. Let the rows of A be $r_1, \dots, r_k$. Their span is unchanged if we interchange two of them or multiply one of them by a nonzero scalar. If we replace the row r_i by $r_i + cr_j$ with $j \neq i$, then the span is unchanged since

$$a_i r_i + a_j r_j = a_i(r_i + cr_j) + (a_j - a_i c)r_j$$

shows that any finite linear combination of the old rows is a finite linear combination of the new rows and since

$$b_i(r_i + cr_j) + b_j r_j = b_i r_i + (b_i c + b_j)r_j$$

shows the reverse. □

Theorem 2.8. If A in $M_{kn}(\mathbb{F})$ has reduced row-echelon form R, then

$$\begin{aligned}\dim(\text{row space}(A)) &= \dim(\text{row space}(R)) \\ &= \#(\text{nonzero rows of } R) = \#(\text{corner variables of } R)\end{aligned}$$

and

$$\dim(\text{null space}(A)) = \dim(\text{null space}(R)) = \#(\text{independent variables of } R).$$

PROOF. The first equality in the first conclusion is immediate from Proposition 2.7, and the last equality of that conclusion is known from the method of row reduction. To see the middle equality, we need to see that the nonzero rows of R are linearly independent. Let these rows be $r_1, \dots, r_t$. For each i with $1 \leq i \leq t$, the index of the first nonzero entry of r_i was denoted by $j(i)$ in Section I.5. That entry has to be 1, and the other rows have to be 0 in that entry, by definition of reduced row-echelon form. If a finite linear combination $c_1 r_1 + \cdots + c_t r_t$ is 0, then inspection of the $j(i)^{\text{th}}$ entry yields the equality $c_i = 0$, and thus we conclude that all the coefficients are 0. This proves the desired linear independence.

The first equality in the second conclusion is by the solution procedure for homogeneous systems of equations in Section I.5; the set of solutions is unchanged by each row operation. To see the second equality, we recall that the form of the solution is as a finite linear combination of specific vectors, the coefficients being the independent variables. What the second equality is asserting is that these vectors form a basis of the space of solutions. We are thus to prove that they are linearly independent. Let the independent variables be certain x_j's, and let the corresponding vectors be v_j's. Then we know that the vector v_j has j^{th} entry 1 and that all the other vectors have j^{th} entry 0. If a finite linear combination of the vectors is 0, then examination of the j^{th} entry shows that the j^{th} coefficient is 0. The result follows. □

Corollary 2.9. If A is in $M_{kn}(\mathbb{F})$, then

$$\dim(\text{row space}(A)) + \dim(\text{null space}(A)) = \#(\text{columns of } A) = n.$$

PROOF. We add the two formulas in Theorem 2.8 and see that

$$\dim(\text{row space}(A)) + \dim(\text{null space}(A))$$

equals the sum $\#(\text{corner variables of } R) + \#(\text{independent variables of } R)$. Since all variables are corner variables or independent variables, this sum is n, and the result follows. □

Corollary 2.10. If A is in $M_{kn}(\mathbb{F})$, then

$$\dim(\text{row space}(A)) = \dim(\text{column space}(A)).$$

REMARK. The common value of the dimension of the row space of A and the dimension of the column space of A is called the **rank** of A. Some authors use the separate terms "row rank" and "column rank" for the two sides, and then the result is that these integers are equal.

PROOF. This follows by comparing Theorem 2.6 and Corollary 2.9. □

Although the above results may seem to have an abstract sound at first, methods of calculation for all the objects in question have quietly been carried along in the proofs, with everything rooted in the method of row reduction. All the proofs have in effect already been given that these methods of calculation do what they are supposed to do. If A is in $M_{kn}(\mathbb{F})$, the **transpose** of A, denoted by A^t, is the member of $M_{nk}(\mathbb{F})$ with entries $(A^t)_{ij} = A_{ji}$. In particular, the transpose of a row vector is a column vector, and vice versa.

METHODS OF CALCULATION.

(1) Basis of the row space of A. Row reduce A, and use the nonzero rows of the reduced row-echelon form.

(2) Basis of the column space of A. Transpose A, compute a basis of the row space of A^t by Method 1, and transpose the resulting row vectors into column vectors.

(3) Basis of the null space of A. Use the solution procedure for $Av = 0$ given in Section I.5. The set of solutions is given as all finite linear combinations of certain column vectors, the coefficients being the independent variables. The column vectors that are obtained form a basis of the null space.

(4) Basis of the linear span of the column vectors $v_1, \ldots, v_n$. Arrange the columns into a matrix A. Then the linear span is the column space of A, and a basis can be determined by Method 2.

(5) Extension of a linearly independent set $\{v_1, \dots, v_r\}$ of column vectors in $\mathbb{F}^n$ to a basis of $\mathbb{F}^n$. Arrange the columns into a matrix, transpose, and row reduce. Adjoin additional row vectors, one for each independent variable, as follows: if x_j is an independent variable, then the row vector corresponding to x_j is to be 1 in the j^{th} entry and 0 elsewhere. Transpose these additional row vectors so that they become column vectors, and these are vectors that may be adjoined to obtain a basis.

(6) Shrinking of a set $\{v_1, \dots, v_r\}$ of column vectors to a subset that is a basis for the linear span of $\{v_1, \dots, v_r\}$. For each i with $0 \le i \le r$, compute $d_i = \dim(\operatorname{span}\{v_1, \dots, v_r\})$. Retain v_i for $i \ge 0$ if $d_{i-1} < d_i$, and discard v_i otherwise.

3. Linear Maps

In this section we discuss linear maps, first in the setting of functions from $\mathbb{F}^n$ to $\mathbb{F}^k$ and then in the setting of functions between two vector spaces over $\mathbb{F}$. Much of the discussion will center on making computations for such functions by means of matrices.

We have seen that any k-by-n matrix A defines a function $L : \mathbb{F}^n$ to $\mathbb{F}^k$ by $L(v) = Av$ and that this function satisfies

$$\begin{aligned} L(u+v) &= L(u) + L(v), \\ L(cv) &= cL(v), \end{aligned}$$

for all u and v in $\mathbb{F}^n$ and all scalars c. A function $L : \mathbb{F}^n \to \mathbb{F}^k$ satisfying these two conditions is said to be **linear**, or $\mathbb{F}$ **linear** if the scalars need emphasizing. Traditional names for such functions are **linear maps**, **linear mappings**, and **linear transformations**.[3] Thus matrices yield linear maps. Here is a converse.

Proposition 2.11. If $L : \mathbb{F}^n \to \mathbb{F}^k$ is a linear map, then there exists a unique k-by-n matrix A such that $L(v) = Av$ for all v in $\mathbb{F}^n$.

REMARK. The proof will show how to obtain the matrix A.

PROOF. For $1 \le j \le n$, let e_j be the j^{th} standard basis vector of $\mathbb{F}^n$, having 1 in its j^{th} entry and 0's elsewhere, and let the j^{th} column of A be the k-dimensional column vector $L(e_j)$. If v is the column vector $(c_1, c_2, \dots, c_n)$, then

$$\begin{aligned} L(v) &= L\big(\textstyle\sum_{j=1}^n c_j e_j\big) = \textstyle\sum_{j=1}^n L(c_j e_j) \\ &= \textstyle\sum_{j=1}^n c_j L(e_j) = \textstyle\sum_{j=1}^n c_j (j^{\text{th}} \text{ column of } A). \end{aligned}$$

[3]The term **linear function** is particularly appropriate when the emphasis is on the fact that a certain function is linear. The term **linear operator** is used also, particularly when the context has something to do with analysis.

If $L(v)_i$ denotes the i^{th} entry of the column vector $L(v)$, this equality says that

$$L(v)_i = \textstyle\sum_{j=1}^n c_j A_{ij}.$$

The right side is the i^{th} entry of Av, and hence $L(v) = Av$. This proves existence. For uniqueness we observe from the formula $L(e_j) = Ae_j$ that the j^{th} column of A has to be $L(e_j)$ for each j, and therefore A is unique. $\square$

In the special case of linear maps from $\mathbb{F}^n$ to $\mathbb{F}^k$, the proof shows that two linear maps that agree on the members of the standard basis are equal on all vectors. We shall give a generalization of this fact as Proposition 2.13 below.

EXAMPLE 1. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation about the origin counterclockwise through the angle θ. Taking L to be defined geometrically, one finds from the parallelogram rule for addition of vectors that L is linear. Computation shows that $L\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and that $L\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$. Applying Proposition 2.11 and the prescription for forming the matrix A given in the proof of the proposition, we see that $L(v) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} v$ for all v in $\mathbb{R}^2$.

We can add two linear maps $L : \mathbb{F}^n \to \mathbb{F}^k$ and $M : \mathbb{F}^n \to \mathbb{F}^k$ by adding their values at corresponding points: $(L + M)(v) = L(v) + M(v)$. In addition, we can multiply a linear map by a scalar by multiplying its values. Then $L + M$ and cL are linear, and it follows that the set of linear maps from $\mathbb{F}^n$ to $\mathbb{F}^k$ is a vector subspace of the vector space of all functions from $\mathbb{F}^n$ to $\mathbb{F}^k$, hence is itself a vector space. The customary notation for this vector space is $\mathrm{Hom}_{\mathbb{F}}(\mathbb{F}^n, \mathbb{F}^k)$; the symbol Hom refers to the validity of the rule $L(u + v) = L(u) + L(v)$, and the subscript $\mathbb{F}$ refers to the validity of the additional rule $L(cv) = cL(v)$ for all c in $\mathbb{F}$.

If L corresponds to the matrix A and M corresponds to the matrix B, then $L + M$ corresponds to $A + B$ and cL corresponds to cA. The next proposition shows that composition of linear maps corresponds to multiplication of matrices.

Proposition 2.12. Let $L : \mathbb{F}^n \to \mathbb{F}^m$ be the linear map corresponding to an m-by-n matrix A, and let $M : \mathbb{F}^m \to \mathbb{F}^k$ be the linear map corresponding to a k-by-m matrix B. Then the composite function $M \circ L : \mathbb{F}^n \to \mathbb{F}^k$ is linear, and it corresponds to the k-by-n matrix BA.

PROOF. The function $M \circ L$ satisfies $(M \circ L)(u + v) = M(L(u + v)) = M(Lu + Lv) = M(Lu) + M(Lv) = (M \circ L)(u) + (M \circ L)(v)$, and similarly it satisfies $(M \circ L)(cv) = c(M \circ L)(v)$. Therefore it is linear. The correspondence of linear maps to matrices and the associativity of matrix multiplication together give $(M \circ L)(v) = M(L(v)) = (B)(Lv) = B(Av) = (BA)v$, and therefore $M \circ L$ corresponds to BA. $\square$

Now let us enlarge the setting for our discussion, treating arbitrary linear maps $L : U \to V$ between vector spaces over $\mathbb{F}$. We say that $L : U \to V$ is **linear**, or $\mathbb{F}$ **linear**, if

$$L(u + v) = L(u) + L(v),$$
$$L(cv) = cL(v),$$

for all u and v in U and all scalars c. As with the special case that $U = \mathbb{F}^n$ and $V = \mathbb{F}^k$, linear functions are called **linear maps**, **linear mappings**, and **linear transformations**. The set of all linear maps $L : U \to V$ is a vector space over $\mathbb{F}$ and is denoted by $\mathrm{Hom}_{\mathbb{F}}(U, V)$. The following result is fundamental in working with linear maps.

Proposition 2.13. Let U and V be vector spaces over $\mathbb{F}$, and let Γ be a basis of U. Then to each function $\ell : \Gamma \to V$ corresponds one and only one linear map $L : U \to V$ whose restriction to Γ has $L\big|_\Gamma = \ell$.

REMARK. We refer to L as the **linear extension** of ℓ.

PROOF. Suppose that $\ell : \Gamma \to V$ is given. Since Γ is a basis of U, each element of U has a unique expansion as a finite linear combination of members of Γ. Say that $u = c_{\alpha_1}u_{\alpha_1} + \cdots + c_{\alpha_r}u_{\alpha_r}$. Then the requirement of linearity on L forces $L(u) = L(c_{\alpha_1}u_{\alpha_1} + \cdots + c_{\alpha_r}u_{\alpha_r}) = c_{\alpha_1}L(u_{\alpha_1}) + \cdots + c_{\alpha_r}L(u_{\alpha_r})$, and therefore L is uniquely determined. For existence, define L by this formula. Expanding u and v in this way, we readily see that $L(u + v) = L(u) + L(v)$ and $L(cu) = cL(u)$. Therefore ℓ has a linear extension. $\square$

The definition of linearity and the proposition just proved make sense even if U and V are infinite-dimensional, but our objective for now will be to understand linear maps in terms of matrices. Thus, until further notice at a point later in this section, we shall assume that U and V are finite-dimensional. Remarks about the infinite-dimensional case appear in Section 9.

Since U and V are arbitrary finite-dimensional vector spaces, we no longer have standard bases at hand, and thus we have no immediate way to associate a matrix to a linear map $L : U \to V$. What we therefore do is fix *arbitrary bases* of U and V and work with them. It will be important to have an enumeration of each of these bases, and we therefore let

$$\Gamma = (u_1, \dots, u_n)$$

and

$$\Delta = (v_1, \dots, v_k)$$

be *ordered* bases of U and V, respectively.[4] If a member u of U may be expanded

[4] The notation $(u_1, \dots, u_n)$ for an ordered basis, with each u_j equal to a vector, is not to be confused with the condensed notation $(c_1, \dots, c_n)$ for a single column vector, with each c_j equal to a scalar.

in terms of Γ as $u = c_1u_1 + \cdots + c_nu_n$, we write

$$\begin{pmatrix} u \\ \Gamma \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

calling this the column vector expressing u in the ordered basis Γ. Using our linear map $L : U \to V$, let us define a k-by-n matrix $\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}$ by requiring that the

$$j^{\text{th}} \text{ column of } \begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix} \qquad \text{be} \qquad \begin{pmatrix} L(u_j) \\ \Delta \end{pmatrix}.$$

The positions in which the ordered bases Δ and Γ are listed in the notation is important here; the range basis is to the left of the domain basis.[5]

EXAMPLE 2. Let V be the space of all complex-valued solutions on $\mathbb{R}$ of the differential equation $y''(t) = y(t)$. Then V is a vector subspace of functions, hence is a vector space in its own right. It is known that V is 2-dimensional with solutions $c_1e^t + c_2e^{-t}$. If $y(t)$ is a solution, then differentiation of the equation shows that $y'(t)$ is another solution. In other words, the derivative operator d/dt is a linear map from V to itself. One ordered basis of V is $\Gamma = (e^t, e^{-t})$, and another is $\Delta = (\cosh t, \sinh t)$, where $\cosh t = \frac{1}{2}(e^t + e^{-t})$ and $\sinh t = \frac{1}{2}(e^t - e^{-t})$. To find $\begin{pmatrix} d/dt \\ \Delta\Gamma \end{pmatrix}$, we need to express $(d/dt)(e^t)$ and $(d/dt)(e^{-t})$ in terms of $\cosh t$ and $\sinh t$. We have

$$\begin{pmatrix} (d/dt)(e^t) \\ \Delta \end{pmatrix} = \begin{pmatrix} e^t \\ \Delta \end{pmatrix} = \begin{pmatrix} \cosh t + \sinh t \\ \Delta \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and $$\begin{pmatrix} (d/dt)(e^{-t}) \\ \Delta \end{pmatrix} = \begin{pmatrix} -e^{-t} \\ \Delta \end{pmatrix} = \begin{pmatrix} -\cosh t + \sinh t \\ \Delta \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Therefore $\begin{pmatrix} d/dt \\ \Delta\Gamma \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

Theorem 2.14. If $L : U \to V$ is a linear map between finite-dimensional vector spaces over $\mathbb{F}$ and if Γ and Δ are ordered bases of U and V, respectively, then

$$\begin{pmatrix} L(u) \\ \Delta \end{pmatrix} = \begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix} \begin{pmatrix} u \\ \Gamma \end{pmatrix}$$

for all u in U.

[5]This order occurs in a number of analogous situations in mathematics and has the effect of keeping the notation reasonably consistent with the notation for composition of functions.

PROOF. The two sides of the identity in question are linear in u, and Proposition 2.13 shows that it is enough to prove the identity for the members u of some ordered basis of U. We choose Γ as this ordered basis. For the basis vector u equal to the j^{th} member u_j of Γ, use of the definition shows that $\begin{pmatrix} u_j \\ \Gamma \end{pmatrix}$ is the column vector e_j that is 1 in the j^{th} entry and is 0 elsewhere. The product $\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix} e_j$ is the j^{th} column of $\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}$, which was defined to be $\begin{pmatrix} L(u_j) \\ \Delta \end{pmatrix}$. Thus the identity in question is valid for u_j, and the theorem follows. $\square$

If we take into account Proposition 2.13, saying that linear maps on U arise uniquely from arbitrary functions on a basis of U, then Theorem 2.14 supplies a one-one correspondence of linear maps L from U to V with matrices A of the appropriate size, once we fix ordered bases in the domain and range. The correspondence is $L \leftrightarrow \begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}$.

As in the special case with linear maps between spaces of column vectors, this correspondence respects addition and scalar multiplication. Theorem 2.14 implies that under this correspondence, the image of L corresponds to the column space of A. It implies also that the vector subspace of the domain U with $L(u) = 0$, which is called the **kernel** of L and is sometimes denoted by $\ker L$, corresponds to the null space of A. The kernel of L has the important property that

$$\text{the linear map } L \text{ is one-one if and only if } \ker L = 0.$$

Another important property comes from this association of kernel with null space and of image with column space. Namely, we apply Theorem 2.6, and we obtain the following corollary.

Corollary 2.15. If $L : U \to V$ is a linear map between finite-dimensional vector spaces over $\mathbb{F}$, then

$$\dim(\text{domain}(L)) = \dim(\text{kernel}(L)) + \dim(\text{image}(L)).$$

The next result says that composition corresponds to matrix multiplication under the correspondence of Theorem 2.14.

Theorem 2.16. Let $L : U \to V$ and $M : V \to W$ be linear maps between finite-dimensional vector spaces, and let Γ, Δ, and Ω be ordered bases of U, V, and W. Then the composition ML is linear, and the corresponding matrix is given by

$$\begin{pmatrix} ML \\ \Omega\Gamma \end{pmatrix} = \begin{pmatrix} M \\ \Omega\Delta \end{pmatrix} \begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}.$$

PROOF. If u is in U, three applications of Theorem 2.14 and one application of associativity of matrix multiplication give

$$\begin{aligned}\begin{pmatrix} ML \\ \Omega\Gamma \end{pmatrix}\begin{pmatrix} u \\ \Gamma \end{pmatrix} &= \begin{pmatrix} ML(u) \\ \Omega \end{pmatrix} = \begin{pmatrix} M \\ \Omega\Delta \end{pmatrix}\begin{pmatrix} L(u) \\ \Delta \end{pmatrix} \\ &= \begin{pmatrix} M \\ \Omega\Delta \end{pmatrix}\left[\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}\begin{pmatrix} u \\ \Gamma \end{pmatrix}\right] = \left[\begin{pmatrix} M \\ \Omega\Delta \end{pmatrix}\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}\right]\begin{pmatrix} u \\ \Gamma \end{pmatrix}.\end{aligned}$$

Taking u to be the j^{th} member of Γ, we see from this equation that the j^{th} column of $\begin{pmatrix} ML \\ \Omega\Gamma \end{pmatrix}$ equals the j^{th} column of $\begin{pmatrix} M \\ \Omega\Delta \end{pmatrix}\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}$. Since j is arbitrary, the theorem follows. $\square$

A computational device that appears at first to be only of theoretical interest and then, when combined with other things, becomes of practical interest, is to change one of the ordered bases in computing the matrix of a linear map. A handy device for this purpose is a change-of-basis matrix $\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}$ since Theorem 2.16 gives $\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix} = \begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}\begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix}$.

EXAMPLE 2, CONTINUED. Let L be d/dt as a linear map carrying the space of solutions of $y''(t) = y(t)$ to itself, with $\Gamma = (e^t, e^{-t})$ and $\Delta = (\cosh t, \sinh t)$ as before. Then $\begin{pmatrix} d/dt \\ \Gamma\Gamma \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Since $e^t = \cosh t + \sinh t$ and $e^{-t} = \cosh t - \sinh t$, $\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ by inspection. The product is $\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix} = \begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}\begin{pmatrix} d/dt \\ \Gamma\Gamma \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, a result we found before with a little more effort by computing matters directly.

Often in practical applications the domain and the range are the same vector space, the domain's ordered basis equals the range's ordered basis, and the matrix of a linear map is known in this ordered basis. The problem is to determine the matrix when the ordered basis is changed in both domain and range—changed in such a way that the ordered bases in the domain and range are the same. This time we use two change-of-basis matrices $\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}$ and $\begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}$, but these are related. Since $\begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix} = \begin{pmatrix} I \\ \Gamma\Gamma \end{pmatrix} = I$, the two matrices are the inverses of one another. Thus, except for matrix algebra, the problem is to compute just one of $\begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}$ and $\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}$.

Normally one of these two matrices can be written down by inspection. For example, if we are working with a linear map from a space of column vectors to itself, one ordered basis of interest is the standard ordered basis Σ. Another ordered basis Δ might be determined by special features of the linear map. In this case the members of Δ are given as column vectors, hence are expressed in terms of Σ. Thus $\begin{pmatrix} I \\ \Sigma\Delta \end{pmatrix}$ can be written by inspection. We shall encounter this situation later in this chapter when we use "eigenvectors" in order to understand linear maps better. Here is an example, but without eigenvectors.

EXAMPLE 1, CONTINUED. We saw that rotation L counterclockwise about the origin in $\mathbb{R}^2$ is given in the standard ordered basis $\Sigma = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$ by $\begin{pmatrix} L \\ \Sigma\Sigma \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Let us compute the matrix of L in the ordered basis $\Delta = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$. The easy change-of-basis matrix to form is $\begin{pmatrix} I \\ \Sigma\Delta \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Hence

$$\begin{pmatrix} L \\ \Delta\Delta \end{pmatrix} = \begin{pmatrix} I \\ \Delta\Sigma \end{pmatrix}\begin{pmatrix} L \\ \Sigma\Sigma \end{pmatrix}\begin{pmatrix} I \\ \Sigma\Delta \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and the problem is reduced to one of matrix algebra.

Our computations have proved the following proposition, which, as we shall see later, motivates much of Chapter V. The matrix C in the statement of the proposition is $\begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}$.

Proposition 2.17. Let $L : V \to V$ be a linear map on a finite-dimensional vector space, and let A be the matrix of L relative to an ordered basis Γ (in domain and range). Then the matrix of L in any other ordered basis Δ is of the form $C^{-1}AC$ for some invertible matrix C depending on Δ.

REMARK. If A is a square matrix, any square matrix of the form $C^{-1}AC$ is said to be **similar** to A. It is immediate that "is similar to" is an equivalence relation.

Now let us return to the setting in which our vector spaces are allowed to be infinite-dimensional. Two vector spaces U and V are said to be **isomorphic** if there is a one-one linear map of U onto V. In this case, the linear map in question is called an **isomorphism**, and one often writes $U \cong V$.

Here is a finite-dimensional example: If U is n-dimensional with an ordered basis Γ and V is k-dimensional with an ordered basis Δ, then $\mathrm{Hom}_{\mathbb{F}}(U, V)$ is isomorphic to $M_{nk}(\mathbb{F})$ by the linear map that carries a member L of $\mathrm{Hom}_{\mathbb{F}}(U, V)$ to the k-by-n matrix $\begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}$.

The relation "is isomorphic to" is an equivalence relation. In fact, it is reflexive since the identity map exhibits U as isomorphic to itself. It is transitive since Theorem 2.16 shows that the composition ML of two linear maps $L : U \to V$ and $M : V \to W$ is linear and since the composition of one-one onto functions is one-one onto. To see that it is symmetric, we need to observe that the inverse function L^{-1} of a one-one onto linear map $L : U \to V$ is linear. To see this linearity, we observe that $L\big(L^{-1}(v_1)+L^{-1}(v_2)\big) = L(L^{-1}(v_1))+L(L^{-1}(v_2)) = v_1 + v_2 = I(v_1 + v_2) = L\big(L^{-1}(v_1 + v_2)\big)$. Since L is one-one,

$$L^{-1}(v_1) + L^{-1}(v_2) = L^{-1}(v_1 + v_2).$$

Similarly the facts that $L(L^{-1}(cv)) = cv = cL(L^{-1}v) = L(c(L^{-1}(v)))$ and that L is one-one imply that

$$L^{-1}(cv) = c(L^{-1}(v)),$$

and hence L^{-1} is linear. Thus "is isomorphic to" is indeed an equivalence relation.

The vector spaces over $\mathbb{F}$ are partitioned, according to the basic result about equivalence relations in Section A2 of the appendix, into equivalence classes. Each member of an equivalence class is isomorphic to all other members of that class and to no member of any other class.

An isomorphism preserves all the vector-space structure of a vector space. Spanning sets are mapped to spanning sets, linearly independent sets are mapped to linearly independent sets, vector subspaces are mapped to vector subspaces, dimensions of subspaces are preserved, and so on. In other words, for all purposes of abstract vector-space theory, isomorphic vector spaces may be regarded as the same. Let us give a condition for isomorphism that might at first seem to trivialize all vector-space theory, reducing it to a count of dimensions, but then let us return to say why this result is not to be considered as so important.

Proposition 2.18. Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic if and only if they have the same dimension.

PROOF. If a vector space U is isomorphic to a vector space V, then the isomorphism carries any basis of U to a basis of V, and hence U and V have the same dimension. Conversely if they have the same dimension, let $(u_1, \dots, u_n)$ be an ordered basis of U, and let $(v_1, \dots, v_n)$ be an ordered basis of V. Define $\ell(u_j) = v_j$ for $1 \le j \le n$, and let $L : U \to V$ be the linear extension of ℓ given by Proposition 2.13. Then L is linear, one-one, and onto, and hence U is isomorphic to V. $\square$

The proposition does not mean that one should necessarily be eager to make the identification of two vector spaces that are isomorphic. An important distinction is the one between "isomorphic" and "isomorphic via a canonically constructed linear map." The isomorphism of linear maps with matrices given by $L \mapsto \begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}$ is canonical since no choices are involved once Γ and Δ have been specified. This is a useful isomorphism because we can track matters down and use the isomorphism to make computations. On the other hand, it is not very useful to say merely that $\mathrm{Hom}_{\mathbb{F}}(U, V)$ and $M_{kn}(\mathbb{F})$ are isomorphic because they have the same dimension.

What tends to happen in practice is that vector spaces in applications come equipped with additional structure—some rigid geometry, or a multiplication operation, or something else. A general vector-space isomorphism has little chance of having any connection to the additional structure and thereby of being very helpful. On the other hand, a concrete isomorphism that is built by taking this additional structure into account may indeed be useful.

In the next section we shall encounter an example of an additional structure that involves neither a rigid geometry nor a multiplication operation. We shall introduce the "dual" V' of a vector space V, and we shall see that V and V' have the same dimension if V is finite-dimensional. But no particular isomorphism of V with V' is singled out as better than other ones, and it is wise not to try to identify these spaces. By contrast, the double dual V'' of V, which too will be constructed in the next section, will be seen to be isomorphic to V in the finite-dimensional case via a linear map $\iota : V \to V''$ that we define explicitly. The function ι is an example of a canonical isomorphism that we might want to exploit.

4. Dual Spaces

Let V be a vector space over $\mathbb{F}$. A **linear functional** on V is a linear map from V into $\mathbb{F}$. The space of all such linear maps, as we saw in Section 3, is a vector space. We denote it by V' and call it the **dual space** of V.

The development of Section 3 tells us right away how to compute the dual space of the space of column vectors $\mathbb{F}^n$. If Σ is the standard ordered basis of $\mathbb{F}^n$ and if 1 denotes the basis of $\mathbb{F}$ consisting of the scalar 1, then we can associate to a linear functional v' on $\mathbb{F}^n$ its matrix

$$\begin{pmatrix} v' \\ 1\Sigma \end{pmatrix} = \begin{pmatrix} v'(e_1) & v'(e_2) & \cdots & v'(e_n) \end{pmatrix},$$

which is an n-dimensional row vector. The operation of v' on a column vector

$v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is given by Theorem 2.14. Namely, $v'(v)$ is a multiple of the scalar 1, and the theorem tells us how to compute this multiple:

$$\begin{pmatrix} v'(v) \\ 1 \end{pmatrix} = \begin{pmatrix} v' \\ 1\Sigma \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} v'(e_1) & v'(e_2) & \cdots & v'(e_n) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Thus the space of all linear functionals on $\mathbb{F}^n$ may be identified with the space of all n-dimensional row vectors, and the effect of the row vector on a column vector is given by matrix multiplication. Since the standard ordered basis of $\mathbb{F}^n$ and the basis 1 of $\mathbb{F}$ are singled out as special, this identification is actually canonical, and it is thus customary to make this identification without further comment.

For a more general vector space V, no natural way of writing down elements of V' comes to mind. Indeed, if a concrete V is given, it can help considerably in understanding V to have an identification of V' that does not involve choices. For example, in real analysis one proves in a suitable infinite-dimensional setting that a (continuous) linear functional on the space of integrable functions is given by integration with a bounded function, and that fact simplifies the handling of the space of integrable functions.

In any event, the canonical identification of linear functionals that we found for $\mathbb{F}^n$ does not work once we pass to a more general finite-dimensional vector space V. To make such an identification in the absence of additional structure, we first fix an ordered basis $(v_1, \ldots, v_n)$ of V. If we do so, then V' is indeed identified with the space of n-dimensional row vectors. The members of V' that correspond to the standard basis of row vectors, i.e., the row vectors that are 1 in one entry and are 0 elsewhere, are of special interest. These are the linear functionals v_i' such that

$$v_i'(v_j) = \delta_{ij},$$

where δ_{ij} is the Kronecker delta. Since these standard row vectors form a basis of the space of row vectors, $(v_1', \ldots, v_n')$ is an ordered basis of V'. If the members of the ordered basis $(v_1, \ldots, v_n)$ are permuted in some way, the members of $(v_1', \ldots, v_n')$ are permuted in the same way. Thus the basis $\{v_1', \ldots, v_n'\}$ depends only on the basis $\{v_1, \ldots, v_n\}$, not on the enumeration.[6] The basis $\{v_1', \ldots, v_n'\}$ is called the **dual basis** of V relative to $\{v_1, \ldots, v_n\}$. A consequence of this discussion is the following result.

Proposition 2.19. If V is a finite-dimensional vector space with dual V', then V' is finite-dimensional with $\dim V' = \dim V$.

[6]Although the enumeration is not important, more structure is present here than simply an association of an unordered basis of V' to an unordered basis of V. Each member of $\{v_1', \ldots, v_n'\}$ is matched to a particular member of $\{v_1, \ldots, v_n\}$, namely the one on which it takes the value 1.

Linear functionals play an important role in working with a vector space. To understand this role, it is helpful to think somewhat geometrically. Imagine the problem of describing a vector subspace of a given vector space. One way of describing it is from the inside, so to speak, by giving a spanning set. In this case we end up by describing the subspace in terms of parameters, the parameters being the scalar coefficients when we say that the subspace is the set of all finite linear combinations of members of the spanning set. Another way of describing the subspace is from the outside, cutting it down by conditions imposed on its elements. These conditions tend to be linear equations, saying that certain linear maps on the elements of the subspace give 0. Typically the subspace is then described as the intersection of the kernels of some set of linear maps. Frequently these linear maps will be scalar-valued, and then we are in a situation of describing the subspace by a set of linear functionals.

We know that every vector subspace of a finite-dimensional vector space V can be described from the inside in this way; we merely give all its members. A statement with more content is that we can describe it with finitely many members; we can do so because we know that every vector subspace of V has a basis.

For linear functionals really to be useful, we would like to know a corresponding fact about describing subspaces from the outside—that every vector subspace U of a finite-dimensional V can be described as the intersection of the kernels of a finite set of linear functionals. To do so is easy. We take a basis of the vector subspace U, say $\{v_1, \dots, v_r\}$, extend it to a basis of V by adjoining vectors $v_{r+1}, \dots, v_n$, and form the dual basis $\{v_1', \dots, v_n'\}$ of V'. The subspace U is then described as the set of all vectors v in V such that $v_j'(v) = 0$ for $r+1 \le j \le n$. The following proposition expresses this fact in ways that are independent of the choice of a basis. It uses the terminology **annihilator** of U, denoted by $\operatorname{Ann}(U)$, for the vector subspace of all members v' of V' with $v'(u) = 0$ for all u in U.

Proposition 2.20. Let V be a finite-dimensional vector space, and let U be a vector subspace of V. Then

(a) $\dim U + \dim \operatorname{Ann}(U) = \dim V$,
(b) every linear functional on U extends to a linear functional on V,
(c) whenever v_0 is a member of V that is not in U, there exists a linear functional on V that is 0 on U and is 1 on v_0.

PROOF. We retain the notation above, writing $\{v_1, \dots, v_r\}$ for a basis of U, $v_{r+1}, \dots, v_n$ for vectors that are adjoined to form a basis of V, and $\{v_1', \dots, v_n'\}$ for the dual basis of V'. For (a), we check that $\{v_{r+1}', \dots, v_n'\}$ is a basis of $\operatorname{Ann}(U)$. It is enough to see that they span $\operatorname{Ann}(U)$. These linear functionals are 0 on every member of the basis $\{v_1, \dots, v_r\}$ of U and hence are in $\operatorname{Ann}(U)$. On the other hand, if v' is a member of $\operatorname{Ann}(U)$, we can certainly write $v' = c_1 v_1' + \cdots + c_n v_n'$

for some scalars $c_1, \ldots, c_n$. Since v' is 0 on U, we must have $v'(v_i) = 0$ for $i \le r$. Since $v'(v_i) = c_i$, we obtain $c_i = 0$ for $i \le r$. Therefore v' is a linear combination of $v'_{r+1}, \ldots, v'_n$, and (a) is proved.

For (b), let us observe that the restrictions $v'_1\big|_U, \ldots, v'_r\big|_U$ form the dual basis of U' relative to the basis $\{v_1, \ldots, v_r\}$ of U. If u' is in U', we can therefore write $u' = c_1 v'_1\big|_U + \cdots + c_r v'_r\big|_U$ for some scalars $c_1, \ldots, c_r$. Then $v' = c_1 v'_1 + \cdots + c_r v'_r$ is the required extension of u' to all of V.

For (c), we use a special choice of basis of V in the argument above. Namely, we still take $\{v_1, \ldots, v_r\}$ to be a basis of U, and then we let $v_{r+1} = v_0$. Finally we adjoin $v_{r+2}, \ldots, v_n$ to obtain a basis $\{v_1, \ldots, v_n\}$ of V. Then v'_{r+1} has the required property. □

If $L : U \to V$ is a linear map between finite-dimensional vector spaces, then the formula

$$(L^t(v'))(u) = v'(L(u)) \qquad \text{for } u \in U \text{ and } v' \in V'$$

defines a linear map $L^t : V' \to U'$. The linear map L^t is called the **contragredient** of L. The matrix of the contragredient of L is the transpose of the matrix of L in the following sense.[7]

Proposition 2.21. Let $L : U \to V$ be a linear map between finite-dimensional vector spaces, let $L^t : V' \to U'$ be its contragredient, let Γ and Δ be respective ordered bases of U and V, and let Γ' and Δ' be their dual ordered bases. Then

$$\begin{pmatrix} L^t \\ \Gamma'\Delta' \end{pmatrix} = \begin{pmatrix} L \\ \Delta\Gamma \end{pmatrix}^t.$$

PROOF. Let $\Gamma = (u_1, \ldots, u_n)$, $\Delta = (v_1, \ldots, v_k)$, $\Gamma' = (u'_1, \ldots, u'_n)$, and $\Delta' = (v'_1, \ldots, v'_k)$. Write B and A for the respective matrices in the formula in question. The equations $L(u_j) = \sum_{i'=1}^k A_{i'j} v_{i'}$ and $L^t(v'_i) = \sum_{j'=1}^n B_{j'i} u'_{j'}$ imply that

$$v'_i(L(u_j)) = v'_i\big(\textstyle\sum_{i'=1}^k A_{i'j} v_{i'}\big) = A_{ij}$$

and

$$L^t(v'_i)(u_j) = \textstyle\sum_{j'=1}^n B_{j'i} u'_{j'}(u_j) = B_{ji}.$$

Therefore $B_{ji} = L^t(v'_i)(u_j) = v'_i(L(u_j)) = A_{ij}$, as required. □

[7] A general principle is involved in the definition of contragredient once we have a definition of dual vector space, and we shall see further examples of this principle in the next two sections and in later chapters: whenever a new systematic construction appears for the objects under study, it is well to look for a corresponding construction with the functions relating these new objects. In language to be introduced near the end of Chapter IV, the context for the construction will be a "category," and the principle says that it is well to see whether the construction is that of a "functor" on the category.

With V finite-dimensional, now consider $V'' = (V')'$, the double dual. In the case that $V = \mathbb{F}^n$, we saw that V' could be viewed as the space of row vectors, and it is reasonable to expect V'' to involve a second transpose and again be the space of column vectors. If so, then V gets identified with V''. In fact, this is true in all cases, and we argue as follows. If v is in V, we can define a member $\iota(v)$ of V'' by

$$\iota(v)(v') = v'(v) \qquad \text{for } v \in V \text{ and } v' \in V'.$$

This definition makes sense whether or not V is finite-dimensional. The function ι is a linear map from V into V'' called the **canonical map** of V into V''. It is independent of any choice of basis.

Proposition 2.22. If V is any finite-dimensional vector space over $\mathbb{F}$, then the canonical map $\iota : V \to V''$ is one-one onto.

REMARKS. In the infinite-dimensional case the canonical map is one-one but it is not onto. The proof that it is one-one uses the fact that V has a basis, but we have deferred the proof of this fact about infinite-dimensional vector spaces to Section 9. Problem 14 at the end of the chapter will give an example of an infinite-dimensional V for which ι does not carry V onto V''. When combined with the first corollary in Section A6 of the appendix, this example shows that ι *never* carries V onto V'' in the infinite-dimensional case.

PROOF. We saw in Section 3 that a linear map ι is one-one if and only if $\ker \iota = 0$. Thus suppose $\iota(v) = 0$. Then $0 = \iota(v)(v') = v'(v)$ for all v'. Arguing by contradiction, suppose $v \neq 0$. Then we can extend $\{v\}$ to a basis of V, and the linear functional v' that is 1 on v and is 0 on the other members of the basis will have $v'(v) \neq 0$, contradiction. We conclude that ι is one-one. By Proposition 2.19 we have

$$\dim V = \dim V' = \dim V''. \tag{$*$}$$

Since ι is one-one, it carries any basis of V to a linearly independent set in V''. This linearly independent set has to be a basis, by Corollary 2.4 and the dimension formula $(*)$. $\square$

5. Quotients of Vector Spaces

This section constructs a vector space V/U out of a vector space V and a vector subspace U. We begin with the example illustrated in Figure 2.1. In the vector space $V = \mathbb{R}^2$, let U be a line through the origin. The lines parallel to U are of the form $v + U = \{v + u \mid u \in U\}$, and we make the set of these lines into a vector space by defining $(v_1 + U) + (v_2 + U) = (v_1 + v_2) + U$ and

$c(v+U) = cv + U$. The figure suggests that if we were to take any other line W through the origin, then W would meet all the lines $v+U$, and the notion of addition of lines $v+U$ would correspond exactly to addition in W. Indeed we can successfully make such a correspondence, but the advantage of introducing the vector space of all lines $v+U$ is that it is canonical, independent of the kind of choice we have to make in selecting W. One example of the utility of having a canonical construction is the ease with which we obtain correspondence of linear maps stated in Proposition 2.25 below. Other examples will appear later.

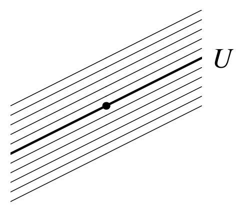


FIGURE 2.1. The vector space of lines $v+U$ in $\mathbb{R}^2$ parallel to a given line U through the origin.

Proposition 2.23. Let V be a vector space over $\mathbb{F}$, and let U be a vector subspace. The relation defined by saying that $v_1 \sim v_2$ if $v_1 - v_2$ is in U is an equivalence relation, and the equivalence classes are all sets of the form $v+U$ with $v \in V$. The set of equivalence classes V/U is a vector space under the definitions

$$\begin{aligned}(v_1+U)+(v_2+U) &= (v_1+v_2)+U,\\ c(v+U) &= cv+U,\end{aligned}$$

and the function $q(v) = v+U$ is linear from V onto V/U with kernel U.

REMARKS. We say that V/U is the **quotient space** of V by U. The linear map $q(v) = v+U$ is called the **quotient map** of V onto V/U.

PROOF. The properties of an equivalence relation are established as follows:

$$\begin{aligned}
v_1 \sim v_1 \qquad & \text{because } 0 \text{ is in } U,\\
v_1 \sim v_2 \text{ implies } v_2 \sim v_1 \qquad & \text{because } U \text{ is closed under negatives,}\\
v_1 \sim v_2 \text{ and } v_2 \sim v_3 \qquad &\\
\text{together imply } v_1 \sim v_3 \qquad & \text{because } U \text{ is closed under addition.}
\end{aligned}$$

Thus we have equivalence classes. The class of v_1 consists of all vectors v_2 such that $v_2 - v_1$ is in U, hence consists of all vectors in $v_1 + U$. Thus the equivalence classes are indeed the sets $v+U$.

Let us check that addition and scalar multiplication, as given in the statement of the proposition, are well defined. For addition let $v_1 \sim w_1$ and $v_2 \sim w_2$. Then $v_1 - w_1$ and $v_2 - w_2$ are in U. Since U is a vector subspace, the sum $(v_1 - w_1) + (v_2 - w_2) = (v_1 + v_2) - (w_1 + w_2)$ is in U. Thus $v_1 + v_2 \sim w_1 + w_2$, and addition is well defined. For scalar multiplication let $v \sim w$, and let a scalar c be given. Then $v - w$ is in U, and $c(v - w) = cv - cw$ is in U since U is a vector subspace. Hence $cv \sim cw$, and scalar multiplication is well defined.

The vector-space properties of V/U are consequences of the properties for V. To illustrate, consider associativity of addition. The argument in this case is that

$$\begin{aligned}
((v_1 + U) + (v_2 + U)) + (v_3 + U) &= ((v_1 + v_2) + U) + (v_3 + U) \\
&= ((v_1 + v_2) + v_3) + U = (v_1 + (v_2 + v_3)) + U \\
&= (v_1 + U) + ((v_2 + v_3) + U) = (v_1 + U) + ((v_2 + U) + (v_3 + U)).
\end{aligned}$$

Finally the quotient map $q : V \to V/U$ given by $q(v) = v + U$ is certainly linear. Its kernel is $\{v \mid v + U = 0 + U\}$, and this equals $\{v \mid v \in U\}$, as asserted. The map q is onto V/U since $v + U = q(v)$. $\square$

Corollary 2.24. If V is a vector space over $\mathbb{F}$ and U is a vector subspace, then

(a) $\dim V = \dim U + \dim(V/U)$,

(b) the subspace U is the kernel of some linear map defined on V.

REMARK. The first conclusion is valid even when all the spaces are not finite-dimensional. For current purposes it is sufficient to regard $\dim V$ as $+\infty$ if V is infinite-dimensional; the sum of $+\infty$ and any dimension as $+\infty$.

PROOF. Let q be the quotient map. The linear map q meets the conditions of (b). For (a), take a basis of U and extend to a basis of V. Then the images under q of the additional vectors form a basis of V/U. $\square$

Quotients of vector spaces allow for the factorization of certain linear maps, as indicated in Proposition 2.25 and Figure 2.2.

Proposition 2.25. Let $L : V \to W$ be a linear map between vector spaces over $\mathbb{F}$, let $U_0 = \ker L$, let U be a vector subspace of V contained in U_0, and let $q : V \to V/U$ be the quotient map. Then there exists a linear map $\overline{L} : V/U \to W$ such that $L = \overline{L}q$. It has the same image as L, and $\ker \overline{L} = \{u_0 + U \mid u_0 \in U_0\}$.

$$\begin{tikzcd}
V \arrow[r, "L"] \arrow[d, "q"'] & W \\
V/U \arrow[ur, dashed, "\overline{L}"'] &
\end{tikzcd}$$

FIGURE 2.2. Factorization of linear maps via a quotient of vector spaces.

REMARK. One says that L **factors through** V/U or **descends** to V/U.

PROOF. The definition of $\overline{L}$ has to be $\overline{L}(v+U) = L(v)$. This forces $\overline{L}q = L$, and $\overline{L}$ will have to be linear. What needs proof is that $\overline{L}$ is well defined. Thus suppose $v_1 \sim v_2$. We are to prove that $\overline{L}(v_1 + U) = \overline{L}(v_2 + U)$, i.e., that $L(v_1) = L(v_2)$. Now $v_1 - v_2$ is in $U \subseteq U_0$, and hence $L(v_1 - v_2) = 0$. Then $L(v_1) = L(v_1 - v_2) + L(v_2) = L(v_2)$, as required. This proves that $\overline{L}$ is well defined, and the conclusions about the image and the kernel of $\overline{L}$ are immediate from the definition. $\square$

Corollary 2.26. Let $L : V \to W$ be a linear map between vector spaces over $\mathbb{F}$, and suppose that L is onto W and has kernel U. Then V/U is canonically isomorphic to W.

PROOF. Take $U = U_0$ in Proposition 2.25, and form $\overline{L} : V/U \to W$ with $L = \overline{L}q$. The proposition shows that $\overline{L}$ is onto W and has trivial kernel, i.e., the 0 element of V/U. Having trivial kernel, $\overline{L}$ is one-one. $\square$

Theorem 2.27 (First Isomorphism Theorem). Let $L : V \to W$ be a linear map between vector spaces over $\mathbb{F}$, and suppose that L is onto W and has kernel U. Then the map $S \mapsto L(S)$ gives a one-one correspondence between

(a) the vector subspaces S of V containing U and
(b) the vector subspaces of W.

REMARK. As in Section A1 of the appendix, we write $L(S)$ and $L^{-1}(T)$ to indicate the direct and inverse images of S and T, respectively.

PROOF. The passage from (a) to (b) is by direct image under L, and the passage from (b) to (a) will be by inverse image under L^{-1}. Certainly the direct image of a vector subspace as in (a) is a vector subspace as in (b). We are to show that the inverse image of a vector subspace as in (b) is a vector subspace as in (a) and that these two procedures invert one another.

For any vector subspace T of W, $L^{-1}(T)$ is a vector subspace of V. In fact, if v_1 and v_2 are in $L^{-1}(T)$, we can write $L(v_1) = t_1$ and $L(v_2) = t_2$ with t_1 and t_2 in T. Then the equations $L(v_1 + v_2) = t_1 + t_2$ and $L(cv_1) = cL(v_1) = ct_1$ show that $v_1 + v_2$ and cv_1 are in $L^{-1}(T)$.

Moreover, the vector subspace $L^{-1}(T)$ contains $L^{-1}(0) = U$. Therefore the inverse image under L of a vector subspace as in (b) is a vector subspace as in (a). Since L is a function, we have $L(L^{-1}(T)) = T$. Thus passing from (b) to (a) and back recovers the vector subspace of W.

If S is a vector subspace of V containing U, we still need to see that $S = L^{-1}(L(S))$. Certainly $S \subseteq L^{-1}(L(S))$. In the reverse direction let v be in $L^{-1}(L(S))$. Then $L(v)$ is in $L(S)$, i.e., $L(v) = L(s)$ for some s in S. Since L

is linear, $L(v - s) = 0$. Thus $v - s$ is in $\ker L = U$, which is contained in S by assumption. Then s and $v - s$ are in S, and hence v is in S. We conclude that $L^{-1}(L(S)) \subseteq S$, and thus passing from (a) to (b) and then back recovers the vector subspace of V containing U. $\square$

If V is a vector space and V_1 and V_2 are vector subspaces, then we write $V_1 + V_2$ for the set $V_1 + V_2$ of all sums $v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$. This is again a vector subspace of V and is called the **sum** of V_1 and V_2. If we have vector subspaces $V_1, \ldots, V_n$, we abbreviate $((\cdots(V_1 + V_2) + V_3) + \cdots + V_n)$ as $V_1 + \cdots + V_n$.

Theorem 2.28 (Second Isomorphism Theorem). Let M and N be vector subspaces of a vector space V over $\mathbb{F}$. Then the map $n + (M \cap N) \mapsto n + M$ is a well-defined canonical vector-space isomorphism

$$N/(M \cap N) \cong (M + N)/M.$$

PROOF. The function $L(n+(M\cap N)) = n+M$ is well defined since $M\cap N \subseteq M$, and L is linear. The domain of L is $\{n + (M \cap N) \mid n \in N\}$, and the kernel is the subset of this where n lies in M as well as N. For this to happen, n must be in $M \cap N$, and thus the kernel is the 0 element of $N/(M \cap N)$. Hence L is one-one.

To see that L is onto $(M+N)/M$, let $(m+n)+M$ be given. Then $n+(M\cap N)$ maps to $n + M$, which equals $(m + n) + M$. Hence L is onto. $\square$

Corollary 2.29. Let M and N be finite-dimensional vector subspaces of a vector space V over $\mathbb{F}$. Then

$$\dim(M + N) + \dim(M \cap N) = \dim M + \dim N.$$

PROOF. Theorem 2.28 and two applications of Corollary 2.24a yield

$$\begin{aligned}\dim(M + N) - \dim M &= \dim((M + N)/M) \\ &= \dim(N/(M \cap N)) = \dim N - \dim(M \cap N),\end{aligned}$$

and the result follows. $\square$

6. Direct Sums and Direct Products of Vector Spaces

In this section we introduce the direct sum and direct product of two or more vector spaces over $\mathbb{F}$. When there are only finitely many such subspaces, these constructions come to the same thing, and we call it "direct sum." We begin with the case that two vector spaces are given.

We define two kinds of direct sums. The **external direct sum** of two vector spaces V_1 and V_2 over $\mathbb{F}$, written $V_1 \oplus V_2$, is a vector space obtained as follows. The underlying set is the set-theoretic product, i.e., the set $V_1 \times V_2$ of ordered pairs (v_1, v_2) with $v_1 \in V_1$ and $v_2 \in V_2$. The operations of addition and scalar multiplication are defined coordinate by coordinate:

$$\begin{aligned}(u_1, u_2) + (v_1, v_2) &= (u_1 + v_1, u_2 + v_2),\\ c(v_1, v_2) &= (cv_1, cv_2),\end{aligned}$$

and it is immediate that $V_1 \oplus V_2$ satisfies the defining properties of a vector space.

If $\{a_i\}$ is a basis of V_1 and $\{b_j\}$ is a basis of V_2, then it follows from the formula $(v_1, v_2) = (v_1, 0) + (0, v_2)$ that $\{(a_i, 0)\} \cup \{(0, b_j)\}$ is a basis of $V_1 \oplus V_2$. Consequently if V_1 and V_2 are finite-dimensional, then $V_1 \oplus V_2$ is finite-dimensional with

$$\dim(V_1 \oplus V_2) = \dim V_1 + \dim V_2.$$

Associated to the construction of the external direct sum of two vector spaces are four linear maps of interest:

$$\begin{array}{llll}\text{two “projections,”} & p_1 : V_1 \oplus V_2 \to V_1 & \text{with } p_1(v_1, v_2) = v_1, \\ & p_2 : V_1 \oplus V_2 \to V_2 & \text{with } p_2(v_1, v_2) = v_2, \\ \text{two “injections,”} & i_1 : V_1 \to V_1 \oplus V_2 & \text{with } i_1(v_1) = (v_1, 0), \\ & i_2 : V_2 \to V_1 \oplus V_2 & \text{with } i_2(v_2) = (0, v_2).\end{array}$$

These have the properties that

$$\begin{aligned} p_r i_s &= \begin{cases} I & \text{on } V_s \text{ if } r = s, \\ 0 & \text{on } V_s \text{ if } r \neq s, \end{cases} \\ i_1 p_1 + i_2 p_2 = I & \qquad \text{on } V_1 \oplus V_2. \end{aligned}$$

The second notion of direct sum captures the idea of recognizing a situation as canonically isomorphic to an external direct sum. This is based on the following proposition.

Proposition 2.30. Let V be a vector space over $\mathbb{F}$, and let V_1 and V_2 be vector subspaces of V. Then the following conditions are equivalent:

(a) every member v of V decomposes uniquely as $v = v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$,
(b) $V_1 + V_2 = V$ and $V_1 \cap V_2 = 0$,
(c) the function from the external direct sum $V_1 \oplus V_2$ to V given by $(v_1, v_2) \mapsto v_1 + v_2$ is an isomorphism of vector spaces.

REMARKS.

(1) If V is a vector space with vector subspaces V_1 and V_2 satisfying the equivalent conditions of Proposition 2.30, then we say that V is the **internal direct sum** of V_1 and V_2. It is customary to write $V = V_1 \oplus V_2$ in this case even though what we have is a canonical isomorphism of the two sides, not an equality.

(2) The dimension formula

$$\dim(V_1 \oplus V_2) = \dim V_1 + \dim V_2$$

for an internal direct sum follows, on the one hand, from the corresponding formula for external direct sums; it follows, on the other hand, by using (b) and Corollary 2.29.

(3) In the proposition it is possible to establish a fourth equivalent condition as follows: there exist linear maps $p_1 : V \to V$, $p_2 : V \to V$, $i_1 : \text{image}\, p_1 \to V$, and $i_2 : \text{image}\, p_2 \to V$ such that

- $p_r i_s p_s$ equals p_r if $r = s$ and equals 0 if $r \neq s$,
- $i_1 p_1 + i_2 p_2 = I$, and
- $V_1 = \text{image}\, i_1 p_1$ and $V_2 = \text{image}\, i_2 p_2$.

PROOF. If (a) holds, then the existence of the decomposition $v = v_1 + v_2$ shows that $V_1 + V_2 = V$. If v is in $V_1 \cap V_2$, then $0 = v + (-v)$ is a decomposition of the kind in (a), and the uniqueness forces $v = 0$. Therefore $V_1 \cap V_2 = 0$. This proves (b).

The function in (c) is certainly linear. If (b) holds and v is given in V, then the identity $V_1 + V_2 = V$ allows us to decompose v as $v = v_1 + v_2$. This proves that the linear map in (c) is onto. To see that it is one-one, suppose that $v_1 + v_2 = 0$. Then $v_1 = -v_2$ shows that v_1 is in $V_1 \cap V_2$. By (b), this intersection is 0. Therefore $v_1 = v_2 = 0$, and the linear map in (c) is one-one.

If (c) holds, then the fact that the linear map in (c) is onto V proves the existence of the decomposition in (a). For uniqueness, suppose that $v_1 + v_2 = u_1 + u_2$ with u_1 and v_1 in V_1 and with u_2 and v_2 in V_2. Then (u_1, u_2) and (v_1, v_2) have the same image under the linear map in (c). Since the function in (c) is assumed one-one, we conclude that $(u_1, u_2) = (v_1, v_2)$. This proves the uniqueness of the decomposition in (a). $\square$

If $V = V_1 \oplus V_2$ is a direct sum, then we can use the above projections and injections to pass back and forth between linear maps with V_1 and V_2 as domain or range and linear maps with V as domain or range. This passage back and forth is called the **universal mapping property** of $V_1 \oplus V_2$ and will be seen later in this section to characterize $V_1 \oplus V_2$ up to canonical isomorphism. Let us be specific about how this property works.

To arrange for V to be the range, suppose that U is a vector space over $\mathbb{F}$ and that $L_1 : U \to V_1$ and $L_2 : U \to V_2$ are linear maps. Then we can define a linear map $L : U \to V$ by $L = i_1 L_1 + i_2 L_2$, i.e., by

$$L(u) = (i_1 L_1 + i_2 L_2)(u) = (L_1(u), L_2(u)),$$

and we can recover L_1 and L_2 from L by $L_1 = p_1 L$ and $L_2 = p_2 L$.

To arrange for V to be the domain, suppose that W is a vector space over $\mathbb{F}$ and that $M_1 : V_1 \to W$ and $M_2 : V_2 \to W$ are linear maps. Then we can define a linear map $M : V \to W$ by $M = M_1 p_1 + M_2 p_2$, i.e., by

$$M(v_1, v_2) = M_1(v_1) + M_2(v_2),$$

and we can recover M_1 and M_2 from M by $M_1 = M i_1$ and $M_2 = M i_2$.

The notion of direct sum readily extends to the direct sum of n vector spaces over $\mathbb{F}$. The **external direct sum** $V_1 \oplus \cdots \oplus V_n$ is the set of ordered pairs $(v_1, \dots, v_n)$ with each v_j in V_j and with addition and scalar multiplication defined coordinate by coordinate. In the finite-dimensional case we have

$$\dim(V_1 \oplus \cdots \oplus V_n) = \dim V_1 + \cdots + \dim V_n.$$

If $V_1, \dots, V_n$ are given as vector subspaces of a vector space V, then we say that V is the **internal direct sum** of $V_1, \dots, V_n$ if the equivalent conditions of Proposition 2.31 below are satisfied. In this case we write $V = V_1 \oplus \cdots \oplus V_n$ even though once again we really have a canonical isomorphism rather than an equality.

Proposition 2.31. Let V be a vector space over $\mathbb{F}$, and let $V_1, \dots, V_n$ be vector subspaces of V. Then the following conditions are equivalent:

(a) every member v of V decomposes uniquely as $v = v_1 + \cdots + v_n$ with $v_j \in V_j$ for $1 \le j \le n$,

(b) $V_1 + \cdots + V_n = V$ and also $V_j \cap (V_1 + \cdots + V_{j-1} + V_{j+1} + \cdots + V_n) = 0$ for each j with $1 \le j \le n$,

(c) the function from the external direct sum $V_1 \oplus \cdots \oplus V_n$ to V given by $(v_1, \dots, v_n) \mapsto v_1 + \cdots + v_n$ is an isomorphism of vector spaces.

Proposition 2.31 is proved in the same way as Proposition 2.30, and the expected analog of Remark 3 with that proposition is valid as well. Notice that the second condition in (b) is stronger than the condition that $V_i \cap V_j = 0$ for all $i \ne j$. Figure 2.3 illustrates how the condition $V_i \cap V_j = 0$ for all $i \ne j$ can be satisfied even though (b) is not satisfied and even though the vector subspaces do not therefore form a direct sum.

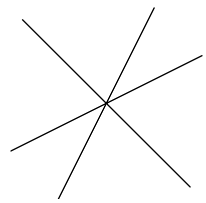


FIGURE 2.3. Three 1-dimensional vector subspaces of $\mathbb{R}^2$ such that each pair has intersection 0.

If $V = V_1 \oplus \cdots \oplus V_n$ is a direct sum, then we can define projections $p_1, \ldots, p_n$ and injections $i_1, \ldots, i_n$ in the expected way, and we again get a universal mapping property. That is, we can pass back and forth between linear maps with $V_1, \ldots, V_n$ as domain or range and linear maps with V as domain or range. The argument given above for $n = 2$ is easily adjusted to handle general n, and we omit the details.

To generalize the above notions to infinitely many vector spaces, there are two quite different ways of proceeding. Let us treat first the external constructions. Let a nonempty collection of vector spaces V_α over $\mathbb{F}$ be given, one for each $\alpha \in A$. The **external direct sum** $\bigoplus_{\alpha \in A} V_\alpha$ is the set of all tuples $\{v_\alpha\}$ in the Cartesian product $\times_{\alpha \in A} V_\alpha$ with all but finitely many v_α equal to 0 and with addition and scalar multiplication defined coordinate by coordinate. For this construction we obtain a basis as the union of embedded bases of the constituent spaces. The **external direct product** $\prod_{\alpha \in A} V_\alpha$ is the set of *all* tuples $\{v_\alpha\}$ in $\times_{\alpha \in A} V_\alpha$, again with addition and scalar multiplication defined coordinate by coordinate. When there are only finitely many factors $V_1, \ldots, V_n$, the external direct product, which manifestly coincides with the external direct sum, is sometimes denoted by $V_1 \times \cdots \times V_n$. For the external direct product when there are infinitely many factors, there is no evident way to obtain a basis of the product from bases of the constituents.

The projections and injections that we defined in the case of finitely many vector spaces are still meaningful here. The universal mapping property is still valid as well, but it splinters into one form for direct sums and another form for direct products. The formulas given above for using linear maps with the V_α's as domain or range to define linear maps with the direct sum or direct product as domain or range may involve sums with infinitely many nonzero terms, and they are not directly usable. Instead, the formulas that continue to make sense are the ones for recovering linear maps with the V_α's as domain or range from linear maps with the direct sum or direct product as domain or range. These turn out to determine the formulas uniquely for the linear maps with the direct sum or direct product as domain or range. In other words, the appropriate universal mapping property uniquely determines the direct sum or direct product up to an isomorphism that respects the relevant projections and injections.

Let us see to the details. We denote typical members of $\prod_{\alpha\in A} V_\alpha$ and $\bigoplus_{\alpha\in A} V_\alpha$ by $\{v_\alpha\}_{\alpha\in A}$, with the understanding that only finitely many v_α can be nonzero in the case of the direct sum. The formulas are

$$p_\beta : \prod_{\alpha\in A} V_\alpha \to V_\beta \qquad \text{with } p_\beta\big(\{v_\alpha\}_{\alpha\in A}\big) = v_\beta,$$

$$i_\beta : V_\beta \to \bigoplus_{\alpha\in A} V_\alpha \qquad \text{with } i_\beta(v_\beta) = \{w_\alpha\}_{\alpha\in A} \text{ and } w_\alpha = \begin{cases} v_\beta & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

If U is a vector space over $\mathbb{F}$ and if a linear map $L_\beta : U \to V_\beta$ is given for each $\beta \in A$, we can obtain a linear map $L : U \to \prod_{\alpha\in A} V_\alpha$ that satisfies $p_\beta L = L_\beta$ for all β. The definition that makes perfectly good sense is

$$L(u) = \{L(u)_\alpha\}_{\alpha\in A} = \{L_\alpha(u)\}_{\alpha\in A}.$$

What does not make sense is to try to express the right side in terms of the injections i_α; we cannot write the right side as $\sum_{\alpha\in A} i_\alpha(L_\alpha(u))$ because infinitely many terms might be nonzero.

If W is a vector space and a linear map $M_\beta : V_\beta \to W$ is given for each β, we can obtain a linear map $M : \bigoplus_{\alpha\in A} V_\alpha \to W$ that satisfies $Mi_\beta = M_\beta$ for all β; the definition that makes perfectly good sense is

$$M\big(\{v_\alpha\}_{\alpha\in A}\big) = \sum_{\alpha\in A} M_\alpha(v_\alpha).$$

The right side is meaningful since only finitely many v_α can be nonzero. It can be misleading to write the formula as $M = \sum_{\alpha\in A} M_\alpha p_\alpha$ because infinitely many of the linear maps $M_\alpha p_\alpha$ can be nonzero functions.

In any event, we have a universal mapping property in both cases—for the direct product with the projections in place and for the direct sum with the injections in place. Let us see that these universal mapping properties characterize direct products and direct sums up to an isomorphism respecting the projections and injections, and that they allow us to define and recognize "internal" direct products and direct sums.

A **direct product** of a set of vector spaces V_α over $\mathbb{F}$ for $\alpha \in A$ consists of a vector space V and a system of linear maps $p_\alpha : V \to V_\alpha$ with the following **universal mapping property**: whenever U is a vector space and $\{L_\alpha\}$ is a system of linear maps $L_\alpha : U \to V_\alpha$, then there exists a unique linear map $L : U \to V$ such that $p_\alpha L = L_\alpha$ for all α. See Figure 2.4. The external direct product establishes existence of a direct product, and Proposition 2.32 below establishes its uniqueness up to an isomorphism of the V's that respects the p_α's. A direct product is said to be **internal** if each V_α is a vector subspace of V and if for each α, the restriction $p_\alpha\big|_{V_\alpha}$ is the identity map on V_α. Because of the uniqueness, this definition of internal direct product is consistent with the earlier one when there are only finitely V_α's.

$$\begin{tikzcd} V_\alpha & U \arrow[l, "L_\alpha"'] \arrow[dl, dashed, "L"] \\ V \arrow[u, "p_\alpha"] & \end{tikzcd}$$

FIGURE 2.4. Universal mapping property of a direct product of vector spaces.

Proposition 2.32. Let A be a nonempty set of vector spaces over $\mathbb{F}$, and let V_α be the vector space corresponding to the member α of A. If $(V, \{p_\alpha\})$ and $(V^*, \{p_\alpha^*\})$ are two direct products of the V_α's, then the linear maps $p_\alpha : V \to V_\alpha$ and $p_\alpha^* : V^* \to V_\alpha$ are onto V_α, there exists a unique linear map $L : V^* \to V$ such that $p_\alpha^* = p_\alpha L$ for all $\alpha \in A$, and L is invertible.

PROOF. In Figure 2.4 let $U = V^*$ and $L_\alpha = p_\alpha^*$. If $L : V^* \to V$ is the linear map produced by the fact that V is a direct product, then we have $p_\alpha L = p_\alpha^*$ for all α. Reversing the roles of V and V^*, we obtain a linear map $L^* : V \to V^*$ with $p_\alpha^* L^* = p_\alpha$ for all α. Therefore $p_\alpha(LL^*) = (p_\alpha L)L^* = p_\alpha^* L^* = p_\alpha$.

In Figure 2.4 we next let $U = V$ and $L_\alpha = p_\alpha$ for all α. Then the identity 1_V on V has the same property $p_\alpha 1_V = p_\alpha$ relative to all p_α that LL^* has, and the uniqueness says that $LL^* = 1_V$. Reversing the roles of V and V^*, we obtain $L^*L = 1_{V^*}$. Therefore L is invertible.

For uniqueness suppose that $\Phi : V^* \to V$ is another linear map with $p_\alpha^* = p_\alpha \Phi$ for all $\alpha \in A$. Then the argument of the previous paragraph shows that $L^*\Phi = 1_{V^*}$. Applying L on the left gives $\Phi = (LL^*)\Phi = L(L^*\Phi) = L1_{V^*} = L$. Thus $\Phi = L$.

Finally we have to show that the α^{th} map of a direct product is onto V_α. It is enough to show that p_α^* is onto V_α. Taking V as the external direct product $\prod_{\alpha \in A} V_\alpha$ with p_α equal to the coordinate mapping, form the invertible linear map $L^* : V \to V^*$ that has just been proved to exist. This satisfies $p_\alpha = p_\alpha^* L^*$ for all $\alpha \in A$. Since p_α is onto V_α, p_α^* must be onto V_α. $\square$

A **direct sum** of a set of vector spaces V_α over $\mathbb{F}$ for $\alpha \in A$ consists of a vector space V and a system of linear maps $i_\alpha : V_\alpha \to V$ with the following **universal mapping property**: whenever W is a vector space and $\{M_\alpha\}$ is a system of linear maps $M_\alpha : V_\alpha \to W$, then there exists a unique linear map $M : V \to W$ such that $Mi_\alpha = M_\alpha$ for all α. See Figure 2.5. The external direct sum establishes existence of a direct sum, and Proposition 2.33 below establishes its uniqueness up to isomorphism of the V's that respects the i_α's. A direct sum is said to be **internal** if each V_α is a vector subspace of V and if for each α, the map i_α is the inclusion map of V_α into V. Because of the uniqueness, this definition of internal direct sum is consistent with the earlier one when there are only finitely V_α's.

$$\begin{tikzcd} V_\alpha \arrow[r, "M_\alpha"] \arrow[d, "i_\alpha"'] & W \\ V \arrow[ur, dashed, "M"'] & \end{tikzcd}$$

FIGURE 2.5. Universal mapping property of a direct sum of vector spaces.

Proposition 2.33. Let A be a nonempty set of vector spaces over $\mathbb{F}$, and let V_α be the vector space corresponding to the member α of A. If $(V, \{i_\alpha\})$ and $(V^*, \{i_\alpha^*\})$ are two direct sums of the V_α's, then the linear maps $i_\alpha : V_\alpha \to V$ and $i_\alpha^* : V_\alpha \to V^*$ are one-one, there exists a unique linear map $M : V \to V^*$ such that $i_\alpha^* = Mi_\alpha$ for all $\alpha \in A$, and M is invertible.

PROOF. In Figure 2.5 let $W = V^*$ and $M_\alpha = i_\alpha^*$. If $M : V \to V^*$ is the linear map produced by the fact that V is a direct sum, then we have $Mi_\alpha = i_\alpha^*$ for all α. Reversing the roles of V and V^*, we obtain a linear map $M^* : V^* \to V$ with $M^*i_\alpha^* = i_\alpha$ for all α. Therefore $(M^*M)i_\alpha = M^*i_\alpha^* = i_\alpha$.

In Figure 2.5 we next let $W = V$ and $M_\alpha = i_\alpha$ for all α. Then the identity 1_V on V has the same property $1_V i_\alpha = i_\alpha$ relative to all i_α that M^*M has, and the uniqueness says that $M^*M = 1_V$. Reversing the roles of V and V^*, we obtain $MM^* = 1_{V^*}$. Therefore M is invertible.

For uniqueness suppose that $\Phi : V \to V^*$ is another linear map with $i_\alpha^* = \Phi i_\alpha$ for all $\alpha \in A$. Then the argument of the previous paragraph shows that $M^*\Phi = 1_V$. Applying M on the left gives $\Phi = (MM^*)\Phi = M(M^*\Phi) = M1_V = M$. Thus $\Phi = M$.

Finally we have to show that the α^{th} map of a direct sum is one-one on V_α. It is enough to show that i_α^* is one-one on V_α. Taking V as the external direct sum $\bigoplus_{s\in S} V_\alpha$ with i_α equal to the embedding mapping, form the invertible linear map $M^* : V^* \to V$ that has just been proved to exist. This satisfies $i_\alpha = M^*i_\alpha^*$ for all $\alpha \in A$. Since i_α is one-one, i_α^* must be one-one. $\square$

7. Determinants

A "determinant" is a certain scalar attached initially to any square matrix and ultimately to any linear map from a finite-dimensional vector space into itself.

The definition is presumably known from high-school algebra in the case of 2-by-2 and 3-by-3 matrices:

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc,$$

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - afh - bdi - ceg.$$

For n-by-n square matrices the determinant function will have the following important properties:

(i) $\det(AB) = \det A \det B$,
(ii) $\det I = 1$,
(iii) $\det A = 0$ if and only if A has no inverse.

Once we have constructed the determinant function with these properties, we can then extend the function to be defined on all linear maps $L : V \to V$ with V finite-dimensional. To do so, we let Γ be any ordered basis of V, and we define $\det L = \det\begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix}$. If Δ is another ordered basis, then

$$\det\begin{pmatrix} L \\ \Delta\Delta \end{pmatrix} = \det\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix} \det\begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix} \det\begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix},$$

and this equals $\det\begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix}$ by (i) since $\begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}$ and $\begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}$ are inverses of each other and since their determinants, by (i) and (ii), are reciprocals. Hence the definition of $\det L$ is independent of the choice of ordered basis, and determinant is well defined on the linear map $L : V \to V$. It is then immediate that the determinant function on linear maps from V into V satisfies (i), (ii), and (iii) above.

Thus it is enough to establish the determinant function on n-by-n matrices. Setting matters up in a useful way involves at least one subtle step, but much of this step has fortunately already been carried out in the discussion of signs of permutations in Section I.4. To proceed, we view det on n-by-n matrices over $\mathbb{F}$ as a function of the n rows of the matrix, rather than the matrix itself. We write V for the vector space $M_{1n}(\mathbb{F})$ of all n-dimensional row vectors. A function $f : V \times \cdots \times V \to \mathbb{F}$ defined on ordered k-tuples of members of V is called a k-**multilinear** functional or k-**linear** functional if it depends linearly on each of the k vector variables when the other $k-1$ vector variables are held fixed. For example,

$$f((a \quad b), (c \quad d)) = ac + b(c+d) + \tfrac{1}{2}ad$$

is a 2-linear functional on $M_{12}(\mathbb{F}) \times M_{12}(\mathbb{F})$. A little more generally and more suggestively,

$$g((a \quad b), (c \quad d)) = \ell_1(a \quad b)\,\ell_2(c \quad d) + \ell_3(a \quad b)\,\ell_4(c \quad d)$$

is a 2-linear functional on $M_{12}(\mathbb{F}) \times M_{12}(\mathbb{F})$ whenever $\ell_1, \ldots, \ell_4$ are linear functionals on $M_{12}(\mathbb{F})$.

Let $\{v_1, \ldots, v_n\}$ be a basis of V. Then a k-multilinear functional as above is determined by its value on all k-tuples of basis vectors $(v_{i_1}, \ldots, v_{i_k})$. (Here $i_1, \ldots, i_k$ are integers between 1 and n.) The reason is that we can fix all but the first variable and expand out the expression by linearity so that only a basis vector remains in each term for the first variable; for each resulting term we can fix all but the second variable and expand out the expression by linearity; and so on. Conversely if we specify arbitrary scalars for the values on each such k-tuple, then we can define a k-multilinear functional assuming those values on the tuples of basis vectors.

A k-multilinear functional f on k-tuples from $M_{1n}(\mathbb{F})$ is said to be **alternating** if f is 0 whenever two of the variables are equal.

EXAMPLE. For $k = 2$ and $n = 2$, we use $\{v_1 = (1 \quad 0),\ v_2 = (0 \quad 1)\}$ as basis. Then a 2-linear multilinear functional f is determined by $f(v_1, v_1)$, $f(v_1, v_2)$, $f(v_2, v_1)$, and $f(v_2, v_2)$. If f is alternating, then $f(v_1, v_1) = f(v_2, v_2) = 0$. But also $f(v_1 + v_2, v_1 + v_2) = 0$, and expansion via 2-multilinearity gives

$$f(v_1, v_1) + f(v_1, v_2) + f(v_2, v_1) + f(v_2, v_2) = 0.$$

We have already seen that the first and last terms on the left side are 0, and thus $f(v_2, v_1) = -f(v_1, v_2)$. Therefore f is completely determined by $f(v_1, v_2)$.

The principle involved in the computation within the example is valid more generally: whenever a multilinear functional f is alternating and two of its arguments are interchanged, then the value of f is multiplied by -1. In fact, let us suppress all variables except for the i^{th} and the j^{th}. Then we have

$$\begin{aligned} 0 &= f(v + w, v + w) = f(v + w, v) + f(v + w, w) \\ &= f(v, v) + f(w, v) + f(v, w) + f(w, w) = f(w, v) + f(v, w). \end{aligned}$$

Theorem 2.34. For $M_{1n}(\mathbb{F})$, the vector space of alternating n-multilinear functionals has dimension 1, and a nonzero such functional has nonzero value on $(e_1^t, \ldots, e_n^t)$, where $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{F}^n$. Let f_0 be the unique such alternating n-multilinear functional taking the value 1 on $(e_1^t, \ldots, e_n^t)$. If a function $\det : M_{nn}(\mathbb{F}) \to \mathbb{F}$ is defined by

$$\det A = f_0(A_{1\cdot}, \ldots, A_{n\cdot})$$

when A has rows $A_{1\cdot}, \ldots, A_{n\cdot}$, then det has the properties that

(a) $\det(AB) = \det A \det B$,
(b) $\det I = 1$,
(c) $\det A = 0$ if and only if A has no inverse,
(d) $\det A = \sum_\sigma (\operatorname{sgn}\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}$, the sum being taken over all permutations σ of $\{1, \ldots, n\}$.

PROOF OF UNIQUENESS. Let f be an alternating n-multilinear functional, and let $\{u_1, \ldots, u_n\}$ be the basis of the space of row vectors defined by $u_i = e_i^t$. Since f is multilinear, f is determined by its values on all n-tuples $(u_{k_1}, \ldots, u_{k_n})$. Since f is alternating, $f(u_{k_1}, \ldots, u_{k_n}) = 0$ unless the u_{k_i} are distinct, i.e., unless $(u_{k_1}, \ldots, u_{k_n})$ is of the form $(u_{\sigma(1)}, \ldots, u_{\sigma(n)})$ for some permutation σ. We have seen that the value of f on an n-tuple of rows is multiplied by -1 if two of the rows are interchanged. Corollary 1.22 and Proposition 1.24b consequently together imply that the value of f on an n-tuple is multiplied by $\operatorname{sgn}\sigma$ if the members of the n-tuple are permuted by σ. Therefore $f(u_{\sigma(1)}, \ldots, u_{\sigma(n)}) = (\operatorname{sgn}\sigma) f(u_1, \ldots, u_n)$, and f is completely determined by its value on $(u_1, \ldots, u_n)$. We conclude that the vector space of alternating n-multilinear functionals has dimension at most 1. $\square$

PROOF OF EXISTENCE. Define $\det A$, and therefore also f_0, by (d). Each term in this definition is the product of n linear functionals, the k^{th} linear functional being applied to the k^{th} argument of f_0, and f_0 is consequently n-multilinear. To see that f_0 is alternating, suppose that the i^{th} and j^{th} rows are equal with $i \neq j$. If τ is the transposition of i and j, then $A_{1\sigma\tau(1)} A_{2\sigma\tau(2)} \cdots A_{n\sigma\tau(n)} = A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}$, and Lemma 1.23 hence shows that

$$(\operatorname{sgn}\sigma\tau) A_{1\sigma\tau(1)} A_{2\sigma\tau(2)} \cdots A_{n\sigma\tau(n)} + (\operatorname{sgn}\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} = 0.$$

Thus if we compute the sum in (d) by grouping pairs of terms, the one for $\sigma\tau$ and the one for σ if $\operatorname{sgn}\sigma = +1$, we see that the whole sum is 0. Thus f_0 is alternating. Finally when A is the identity matrix I, we see that $A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} = 0$ unless σ is the identity permutation, and then the product is 1. Since $\operatorname{sgn} 1 = +1$, $\det I = +1$. We conclude that the vector space of alternating n-multilinear functionals has dimension exactly 1. $\square$

PROOF OF PROPERTIES OF det. Fix an n-by-n matrix B. Since f_0 is alternating n-multilinear, so is $(v_1, \ldots, v_n) \mapsto f_0(v_1 B, \ldots, v_n B)$. The vector space of alternating n-multilinear functionals has been proved to be of dimension 1, and therefore $f_0(v_1 B, \ldots, v_n B) = c(B) f_0(v_1, \ldots, v_n)$ for some scalar $c(B)$. In the notation with det, this equation reads $\det(AB) = c(B)\det A$. Putting $A = I$, we obtain $\det B = c(B)\det I$. Thus $c(B) = \det B$, and (a) follows. We have already proved (b), and (d) was the definition of $\det A$. We are left with (c). If A^{-1}

exists, then (a) and (b) give $\det(A^{-1})\det A = \det I = 1$, and hence $\det A \neq 0$. If A^{-1} does not exist, then Theorem 1.30 and Proposition 1.27c show that the reduced row-echelon form R of A has a row of 0's. We combine Proposition 1.29, conclusion (a), the invertibility of elementary matrices, and the fact that invertible matrices have nonzero determinant, and we see that $\det A$ is the product of $\det R$ and a nonzero scalar. Since det is linear as a function of each row and since R has a row of 0's, $\det R = 0$. Therefore $\det A = 0$. This completes the proof of the theorem. $\square$

The fast procedure for evaluating determinants is to use row reduction, keeping track of what happens. The effect of each kind of row operation on a determinant and the reasons the function det behaves in this way are as follows:

(i) Interchange two rows. This operation multiplies the determinant by -1 because of the alternating property.

(ii) Multiply a row by a nonzero scalar c. This operation multiplies the determinant by c because of the linearity of determinant as a function of that row.

(iii) Replace the i^{th} row by the sum of it and a multiple of the j^{th} row with $j \neq i$. This operation leaves the determinant unchanged. In fact, the matrix whose i^{th} row is replaced by the j^{th} row has determinant 0 by the alternating property, and the rest follows by linearity in the i^{th} row.

As with row reduction the number of steps required to compute a determinant this way is $\leq Cn^3$ in the n-by-n case.

A certain savings of computation is possible as compared with full-fledged row reduction. Namely, we have only to arrange for the reduced matrix to be 0 below the main diagonal, and then the determinant of the reduced matrix will be the product of the diagonal entries, by inspection of the formula in Theorem 2.34d.

EXAMPLE. For the matrix $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$, we have

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \overset{\text{(iii)}}{=} \det\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{pmatrix}$$
$$\overset{\text{(ii)}}{=} -3\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -6 & -11 \end{pmatrix} \overset{\text{(iii)}}{=} -3\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = -3.$$

We conclude this section with a number of formulas for determinants.

Proposition 2.35. If A is an n-by-n square matrix, then $\det A^t = \det A$.

PROOF. Corollary 2.9 says that the row space and the column space of A have the same dimension, and A is invertible if and only if the row space has dimension n. Thus A is invertible if and only if A^t is invertible, and Theorem 2.34c thus shows that $\det A = 0$ if and only if $\det A^t = 0$. Now suppose that $\det A$ and $\det A^t$ are nonzero. Then we can write $A = E_1 \cdots E_r$ with each E_j an elementary matrix of one of the three types. Theorem 2.34a shows that $\det A = \prod_{j=1}^r \det E_j$ and $\det A^t = \prod_{j=1}^r \det E_j^t$, and hence it is enough to prove that $\det E_j = \det E_j^t$ for each j. For E_j of either of the first two types, $E_j = E_j^t$ and there is nothing to prove. For E_j of the third type, we have $\det E_j = \det E_j^t = 1$. The result follows. $\square$

Proposition 2.36 (expansion in cofactors). Let A be an n-by-n matrix, and let $\widehat{A_{ij}}$ be the square matrix of size $n-1$ obtained by deleting the i^{th} row and the j^{th} column. Then

(a) for any j, $\det A = \sum_{i=1}^n (-1)^{i+j} A_{ij} \det \widehat{A_{ij}}$, i.e., $\det A$ may be calculated by "expansion in cofactors" about the j^{th} column,

(b) for any i, $\det A = \sum_{j=1}^n (-1)^{i+j} A_{ij} \det \widehat{A_{ij}}$, i.e., $\det A$ may be calculated by "expansion in cofactors" about the i^{th} row.

REMARKS. If this formula is iterated, we obtain a procedure for evaluating a determinant in about $Cn!$ steps. This procedure amounts to using the formula for $\det A$ in Theorem 2.34d and is ordinarily not of practical use. However, it *is* of theoretical use, and Corollary 2.37 will provide a simple example of a theoretical application.

PROOF. It is enough to prove (a) since (b) then follows by combining (a) and Proposition 2.35. In (a), the right side is 1 when $A = I$, and it is enough by Theorem 2.34 to prove that the right side is alternating and n-multilinear. Each term on the right side is n-multilinear, and hence so is the whole expression. To see that the right side is alternating, suppose that the k^{th} and l^{th} rows are equal with $k < l$. The k^{th} and l^{th} rows are both present in $\widehat{A_{ij}}$ if i is not equal to k or l, and thus each $\det \widehat{A_{ij}}$ is 0 for i not equal to k or l. We are left with showing that

$$(-1)^{k+j} A_{kj} \det \widehat{A_{kj}} + (-1)^{l+j} A_{lj} \det \widehat{A_{lj}} = 0.$$

The two matrices $\widehat{A_{kj}}$ and $\widehat{A_{lj}}$ have the same rows but in a different order. The order is

$$\begin{array}{ll} 1, \dots, k-1, k+1, \dots, l-1, l, l+1, \dots, n & \qquad \text{in the case of } \widehat{A_{kj}}, \\ 1, \dots, k-1, k, k+1, \dots, l-1, l+1, \dots, n & \qquad \text{in the case of } \widehat{A_{lj}}. \end{array}$$

We can transform the first matrix into the second by transposing the index for row l to the left one step at a time until it gets to the k^{th} position. The number of steps is $l-k-1$, and therefore $\det \widehat{A_{lj}} = (-1)^{l-k-1} \det \widehat{A_{kj}}$. Consequently

$$\begin{aligned}(-1)^{k+j} A_{kj} \det \widehat{A_{kj}} &+ (-1)^{l+j} A_{lj} \det \widehat{A_{lj}} \\ &= \big((-1)^{k+j} A_{kj} + (-1)^{2l-k-1+j} A_{lj}\big) \det \widehat{A_{kj}}.\end{aligned}$$

The right side is 0 since $A_{kj} = A_{lj}$, and the proof is complete. $\square$

Corollary 2.37 (Vandermonde matrix and determinant). If $r_1, \dots, r_n$ are scalars, then

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ r_1^2 & r_2^2 & \cdots & r_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_1^{n-1} & r_2^{n-1} & \cdots & r_n^{n-1} \end{pmatrix} = \prod_{j>i} (r_j - r_i).$$

PROOF. We show that the determinant is

$$= \prod_{j>1} (r_j - r_1) \det \begin{pmatrix} 1 & \cdots & 1 \\ r_2 & \cdots & r_n \\ \vdots & \ddots & \vdots \\ r_2^{n-2} & \cdots & r_n^{n-2} \end{pmatrix},$$

and then the result follows by induction. In the given matrix, replace the n^{th} row by the sum of it and $-r_1$ times the $(n-1)^{\text{st}}$ row, then the $(n-1)^{\text{st}}$ row by the sum of it and $-r_1$ times the $(n-2)^{\text{nd}}$ row, and so on. The resulting determinant is

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & r_2 - r_1 & \cdots & r_n - r_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & r_2^{n-2} - r_1 r_2^{n-3} & \cdots & r_n^{n-2} - r_1 r_n^{n-3} \\ 0 & r_2^{n-1} - r_1 r_2^{n-2} & \cdots & r_n^{n-1} - r_1 r_n^{n-2} \end{pmatrix}$$

$$= \det \begin{pmatrix} r_2 - r_1 & \cdots & r_n - r_1 \\ \vdots & \ddots & \vdots \\ r_2^{n-2} - r_1 r_2^{n-3} & \cdots & r_n^{n-2} - r_1 r_n^{n-3} \\ r_2^{n-1} - r_1 r_2^{n-2} & \cdots & r_n^{n-1} - r_1 r_n^{n-2} \end{pmatrix} \qquad \text{by Proposition 2.36a applied with } j = 1$$

$$= (r_2 - r_1) \cdots (r_n - r_1) \det \begin{pmatrix} 1 & \cdots & 1 \\ r_2 & \cdots & r_n \\ \vdots & \ddots & \vdots \\ r_2^{n-2} & \cdots & r_n^{n-2} \end{pmatrix},$$

the last step following by multilinearity of the determinant in the columns (as a consequence of Proposition 2.35 and multilinearity in the rows). $\square$

The **classical adjoint** of the square matrix A, denoted by A^{adj}, is the matrix with entries $A^{\mathrm{adj}}_{ij} = (-1)^{i+j} \det \widehat{A_{ji}}$ with $\widehat{A_{kl}}$ defined as in the statement of Proposition 2.36: $\widehat{A_{kl}}$ is the matrix A with the k^{th} row and l^{th} column deleted.

In the 2-by-2 case, we have $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{\mathrm{adj}} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Thus we have $AA^{\mathrm{adj}} = A^{\mathrm{adj}}A = (\det A)I$ in the 2-by-2 case. Cramer's rule for solving simultaneous linear equations results from the n-by-n generalization of this formula.

Proposition 2.38 (Cramer's rule). If A is an n-by-n matrix, then $AA^{\mathrm{adj}} = A^{\mathrm{adj}}A = (\det A)I$, and thus $\det A \neq 0$ implies $A^{-1} = (\det A)^{-1}A^{\mathrm{adj}}$. Consequently if $\det A \neq 0$, then the unique solution of the simultaneous system $Ax = b$ of n equations in n unknowns, in which $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$, has

$$x_j = \frac{\det B_j}{\det A}$$

with B_j equal to the n-by-n matrix obtained from A by replacing the j^{th} column of A by b.

REMARKS. If we think of the calculation of the determinant of an n-by-n matrix as requiring about n^3 steps, then application of Cramer's rule, at least if done in an unthinking fashion, suggests that solving an invertible system requires about $n^3(n+1)$ steps, i.e., $n+1$ determinants are involved in the explicit solution. Use of row reduction directly to solve the system is more efficient than proceeding this way. Thus Cramer's rule is more important for its theoretical applications than it is for making computations. One simple theoretical application is the observation that each entry of the inverse of a matrix is the quotient of a polynomial function of the entries divided by the determinant.

PROOF. The $(i, j)^{\text{th}}$ entry of $A^{\mathrm{adj}}A$ is

$$(A^{\mathrm{adj}}A)_{ij} = \sum_{k=1}^{n} A^{\mathrm{adj}}_{ik} A_{kj} = \sum_{k=1}^{n} (-1)^{i+k} (\det \widehat{A_{ki}}) A_{kj}.$$

If $i = j$, then expansion in cofactors about the j^{th} column (Proposition 2.36a) identifies the right side as $\det A$. If $i \neq j$, consider the matrix B obtained from A by replacing the i^{th} column of A by the j^{th} column. Then the i^{th} and j^{th} columns of B are equal, and hence $\det B = 0$. Expanding $\det B$ in cofactors about the i^{th} column (Proposition 2.36a), we obtain

$$0 = \det B = \sum_{k=1}^{n} (-1)^{i+k} (\det \widehat{B_{ki}}) B_{ki} = \sum_{k=1}^{n} (-1)^{i+k} (\det \widehat{A_{ki}}) A_{kj}.$$

Thus $AA^{\text{adj}} = (\det A)I$. A similar argument proves that $A^{\text{adj}}A = (\det A)I$.

For the application to $Ax = b$, we multiply both sides on the left by A^{adj} and obtain $(\det A)x = A^{\text{adj}}b$. Hence

$$(\det A)x_j = \sum_{i=1}^{n} (A^{\text{adj}})_{ji} b_i = \sum_{i=1}^{n} (-1)^{i+j} b_i \det \widehat{A_{ij}},$$

and the right side equals $\det B_j$ by expansion in cofactors of $\det B_j$ about the j^{th} column (Proposition 2.36a). $\square$

8. Eigenvectors and Characteristic Polynomials

A vector $v \neq 0$ in $\mathbb{F}^n$ is an **eigenvector** of the n-by-n matrix A if $Av = \lambda v$ for some scalar λ. We call λ the **eigenvalue** associated with v. When λ is an eigenvalue, the vector space of all v with $Av = \lambda v$, i.e., the set consisting of the eigenvectors and the 0 vector, is called the **eigenspace** for λ.

If we think of A as giving a linear map L from $\mathbb{F}^n$ to itself, an eigenvector takes on geometric significance as a vector mapped to a multiple of itself by L. Another geometric way of viewing matters is that the eigenvector yields a 1-dimensional subspace $U = \mathbb{F}v$ that is **invariant**, or **stable**, under L in the sense of satisfying $L(U) \subseteq U$.

Proposition 2.39. An n-by-n matrix A has an eigenvector with eigenvalue λ if and only if $\det(\lambda I - A) = 0$. In this case the eigenspace for λ is the kernel of $\lambda I - A$.

PROOF. We have $Av = \lambda v$ if and only if $(\lambda I - A)v = 0$, if and only if v is in $\ker(\lambda I - A)$. This kernel is nonzero if and only if $\det(\lambda I - A) = 0$. $\square$

With A fixed, the expression $\det(\lambda I - A)$ is a polynomial in λ of degree n and is called the **characteristic polynomial**[8] of A. To see that it is at least a polynomial function of λ, let us expand $\det(\lambda I - A)$ as

$$\det\begin{pmatrix} \lambda - A_{11} & -A_{12} & \cdots & -A_{1n} \\ -A_{21} & \lambda - A_{22} & \cdots & -A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -A_{n1} & -A_{n2} & \cdots & \lambda - A_{nn} \end{pmatrix}$$
$$= \sum_{\sigma} (\operatorname{sgn}\sigma)\operatorname{term}_{1,\sigma(1)} \cdots \operatorname{term}_{n,\sigma(n)}.$$

The term for the permutation $\sigma = 1$ has $\sigma(k) = k$ for every k and gives $\prod_{j=1}^n (\lambda - A_{jj})$. All other σ's have $\sigma(k) = k$ for at most $n-2$ values of k, and λ therefore occurs at most $n-2$ times. Thus the above expression is

$$= \prod_{j=1}^n (\lambda - A_{jj}) + \begin{Bmatrix}\text{other terms with powers} \\ \text{of } \lambda \text{ at most } n-2\end{Bmatrix}$$
$$= \lambda^n - \Big(\sum_{j=1}^n A_{jj}\Big)\lambda^{n-1} + \begin{Bmatrix}\text{terms with powers of} \\ \lambda \text{ from } n-2 \text{ to } 1\end{Bmatrix} + (-1)^n \det A.$$

The constant term is $(-1)^n \det A$ as indicated because it is the value of the polynomial at $\lambda = 0$, which is $\det(-A)$. In any event, we now see that characteristic polynomials are polynomial functions. Starting in Chapter V, we shall treat them as polynomials in one indeterminate in the sense[9] of Section I.3; for now, we are calling the indeterminate λ, but later as our point of view evolves, we shall start calling it X. The negative of the coefficient of λ^{n-1} is the **trace** of A, denoted by $\operatorname{Tr} A$. Thus $\operatorname{Tr} A = \sum_{j=1}^n A_{jj}$. Trace is a linear functional on the vector space $M_{nn}(\mathbb{F})$ of n-by-n matrices.

EXAMPLE 1. For $A = \begin{pmatrix} 4 & 1 \\ -2 & 1 \end{pmatrix}$, the characteristic polynomial is

$$\det(\lambda I - A) = \det\begin{pmatrix} \lambda - 4 & -1 \\ 2 & \lambda - 1 \end{pmatrix}$$
$$= (\lambda - 4)(\lambda - 1) + 2 = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3).$$

[8]Some authors call $\det(A - \lambda I)$ the characteristic polynomial. This is the same polynomial as $\det(\lambda I - A)$ if n is even and is the negative of it if n is odd. The choice made here has the slight advantage of always having leading coefficient 1, which is a handy property in some situations.

[9]In Chapter V we will allow determinants of matrices whose entries are from any "commutative ring with identity," $\mathbb{C}[\lambda]$ being an example. Then we can think of $\det(\lambda I - A)$ directly as involving an indeterminate λ and not initially as a function of a scalar λ.

The roots, and hence the eigenvalues, are $\lambda = 2$ and $\lambda = 3$. The eigenvectors for $\lambda = 2$ are computed by solving $(2I - A)v = 0$. The method of row reduction gives

$$\left(\begin{array}{cc|c} 2-4 & -1 & 0 \\ 2 & 2-1 & 0 \end{array}\right) = \left(\begin{array}{cc|c} -2 & -1 & 0 \\ 2 & 1 & 0 \end{array}\right) \mapsto \left(\begin{array}{cc|c} 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Thus we have $x_1 + \frac{1}{2}x_2 = 0$ and $x_1 = -\frac{1}{2}x_2$. So the eigenvectors for $\lambda = 2$ are the nonzero vectors of the form $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}$. Similarly we find the eigenvectors for $\lambda = 3$ by starting from $(3I - A)v = 0$ and solving. The result is that the eigenvectors for $\lambda = 3$ are the nonzero vectors of the form $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. For this example, there is a basis of eigenvectors.

Corollary 2.40. An n-by-n matrix A has at most n eigenvalues.

PROOF. Since $\det(\lambda I - A)$ is a polynomial of degree n, this follows from Proposition 2.39 and Corollary 1.14. $\square$

It will later be of interest that certain matrices A have a basis of eigenvectors. Such a basis exists for A as in Example 1 but not in general. One thing that can prevent a matrix from having a basis of eigenvectors is the failure of the characteristic polynomial to factor into first-degree factors. Thus, for example, $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has characteristic polynomial $\lambda^2 + 1$, which does not factor into first-degree factors when $\mathbb{F} = \mathbb{R}$. Even when we do have a factorization into first-degree factors, we can still fail to have a basis of eigenvectors, as the following example shows.

EXAMPLE 2. For $A = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$, the characteristic polynomial is given by $\det(\lambda I - A) = \det \begin{pmatrix} \lambda - 1 & 1 \\ 0 & \lambda - 1 \end{pmatrix} = (\lambda - 1)^2$. When we solve for eigenvectors, we get $\left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$, and $x_2 = 0$. Thus $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and we do not have a basis of eigenvectors.

What happens is that the presence of a factor $(\lambda - c)^k$ in the characteristic polynomial ensures the existence of an r-parameter family of eigenvectors for eigenvalue c, with $1 \le r \le k$, but not necessarily with $r = k$. Example 2 shows that r can be strictly less than k. For purposes of deciding whether there is a basis of eigenvectors, the positive result is that the different roots of the characteristic polynomial do not interfere with each other; this is a consequence of the following proposition.

Proposition 2.41. If A is an n-by-n matrix, then eigenvectors for distinct eigenvalues are linearly independent.

REMARK. It follows that if the characteristic polynomial of A has n distinct eigenvalues, then it has a basis of eigenvectors.

PROOF. Let $Av_1 = \lambda_1 v_1, \dots, Av_k = \lambda_k v_k$ with $\lambda_1, \dots, \lambda_k$ distinct, and suppose that

$$c_1 v_1 + \cdots + c_k v_k = 0.$$

Applying A repeatedly gives

$$\begin{aligned} c_1\lambda_1 v_1 + \cdots + c_k\lambda_k v_k &= 0, \\ c_1\lambda_1^2 v_1 + \cdots + c_k\lambda_k^2 v_k &= 0, \\ &\vdots \\ c_1\lambda_1^{k-1} v_1 + \cdots + c_k\lambda_k^{k-1} v_k &= 0. \end{aligned}$$

If the j^{th} entry of v_i is denoted by $v_i^{(j)}$, this system of vector equations says that

$$\begin{pmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_k \\ \vdots & \ddots & \vdots \\ \lambda_1^{k-1} & \cdots & \lambda_k^{k-1} \end{pmatrix} \begin{pmatrix} c_1 v_1^{(j)} \\ \vdots \\ c_k v_k^{(j)} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{for } 1 \le j \le n.$$

The square matrix on the left side is a Vandermonde matrix, which is invertible by Corollary 2.37 since $\lambda_1, \dots, \lambda_k$ are distinct. Therefore $c_i v_i^{(j)} = 0$ for all i and j. Each v_i is nonzero in some entry $v_i^{(j)}$ with j perhaps depending on i, and hence $c_i = 0$. Since all the coefficients c_i have to be 0, $v_1, \dots, v_k$ are linearly independent. □

The theory of eigenvectors and eigenvalues for square matrices allows us to develop a corresponding theory for linear maps $L : V \to V$, where V is an n-dimensional vector space over $\mathbb{F}$. If L is such a function, a vector $v \neq 0$ in V is an **eigenvector** of L if $L(v) = \lambda v$ for some scalar λ. We call λ the **eigenvalue**. When λ is an eigenvalue, the vector space of all v with $L(v) = \lambda v$ is called the **eigenspace** for λ under L. We can compute the eigenvalues and eigenvectors of L by working in any ordered basis Γ of V. The equation $L(v) =$

λv becomes $\begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix}\begin{pmatrix} v \\ \Gamma \end{pmatrix} = \lambda \begin{pmatrix} v \\ \Gamma \end{pmatrix}$ and is satisfied if and only if the column vector $\begin{pmatrix} v \\ \Gamma \end{pmatrix}$ is an eigenvector of the matrix $A = \begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix}$ with eigenvalue λ. Applying Proposition 2.39 and remembering that determinants are well defined on linear maps $L : V \to V$, we see that L has an eigenvector with eigenvalue λ if and only if $\det(\lambda I - L) = 0$ and that in this case the eigenspace is the kernel of $\lambda I - L$.

What happens if we make these computations in a different ordered basis Δ? We know from Proposition 2.17 that the matrices $A = \begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix}$ and $B = \begin{pmatrix} L \\ \Delta\Delta \end{pmatrix}$ are similar, related by $B = C^{-1}AC$, where $C = \begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}$. Computing with A leads to $u = \begin{pmatrix} v \\ \Gamma \end{pmatrix}$ as eigenvector for the eigenvalue λ. The corresponding result for B is that $B(C^{-1}u) = C^{-1}ACC^{-1}u = C^{-1}Au = \lambda C^{-1}u$. Thus $C^{-1}u = \begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix}\begin{pmatrix} v \\ \Gamma \end{pmatrix} = \begin{pmatrix} v \\ \Delta \end{pmatrix}$ is an eigenvector of B with eigenvalue λ, just as it should be.

These considerations about eigenvalues suggest some facts about similar matrices that we can observe more directly without first passing from matrices to linear maps: One is that similar matrices have the same characteristic polynomial. To see this, suppose that $B = C^{-1}AC$; then

$$\begin{aligned} \det(\lambda I - B) &= \det(\lambda I - C^{-1}AC) = \det(C^{-1}(\lambda I - A)C) \\ &= (\det C^{-1}) \det(\lambda I - A)(\det C^{-1}) \\ &= (\det C^{-1})(\det C^{-1}) \det(\lambda I - A) = \det(\lambda I - A). \end{aligned}$$

A second fact is that similar matrices have the same trace. In fact, the trace is the negative of the coefficient of λ^{n-1} in the characteristic polynomial, and the characteristic polynomials are the same.

Because of these considerations we are free in the future to speak of the characteristic polynomial, the eigenvalues, and the trace of a linear map from a finite-dimensional vector space to itself, as well as the determinant, and these notions do not depend on any choice of ordered basis. We can speak unambiguously also of the eigenvectors of such a linear map. For this notion the realization of the eigenvectors in an ordered basis as column vectors depends on the ordered basis, the dependence being given by the formulas two paragraphs before the present one.

One final remark is in order. When the scalars are taken to be the complex numbers $\mathbb{C}$, the Fundamental Theorem of Algebra (Theorem 1.18) is applicable:

every polynomial of degree ≥ 1 has at least one root. When applied to the characteristic polynomial of a square matrix or a linear map from a finite-dimensional vector space to itself, this theorem tells us that the matrix or linear map always has at least one eigenvalue, hence an eigenvector. We shall make serious use of this fact in Chapter III.

9. Bases in the Infinite-Dimensional Case

So far in this chapter, the use of bases has been limited largely to vector spaces having a finite spanning set. In this case we know from Corollary 2.3 that the finite spanning set has a subset that is a basis, any linearly independent set can be extended to a basis, and any two bases have the same finite number of elements. We called such spaces finite-dimensional and defined the dimension of the vector space to be the number of elements in a basis.

The first objective in this section is to prove analogs of these results in the infinite-dimensional case. We shall make use of Zorn's Lemma as in Section A5 of the appendix, as well as the notion of cardinality discussed in Section A6 of the appendix. Once these analogs are in place, we shall examine the various results that we proved about finite-dimensional spaces to see the extent to which they remain valid for infinite-dimensional spaces.

Theorem 2.42. If V is any vector space over $\mathbb{F}$, then

(a) any spanning set in V has a subset that is a basis,
(b) any linearly independent set in V can be extended to a basis,
(c) V has a basis,
(d) any two bases have the same cardinality.

REMARKS. The common cardinality mentioned in (d) is called the **dimension** of the vector space V. In many applications it is enough to use $+\infty$ in place of each infinite cardinal in dimension formulas. This was the attitude conveyed in the remark with Corollary 2.24.

PROOF. For (b), let E be the given linearly independent set, and let $\mathcal{S}$ be the collection of all linearly independent subsets of V that contain E. Partially order $\mathcal{S}$ by inclusion upward. The set $\mathcal{S}$ is nonempty because E is in $\mathcal{S}$. Let $\mathcal{T}$ be a chain in $\mathcal{S}$, and let A be the union of the members of $\mathcal{T}$. We show that A is in $\mathcal{S}$, and then A is certainly an upper bound of $\mathcal{T}$. Because of its definition, A contains E, and we are to prove that A is linearly independent. For A to fail to be linearly independent would mean that there are vectors $v_1, \ldots, v_n$ in A with $c_1v_1 + \cdots + c_nv_n = 0$ for some system of scalars not all 0. Let v_j be in the member A_j of the chain $\mathcal{T}$. Since $A_1 \subseteq A_2$ or $A_2 \subseteq A_1$, v_1 and v_2 are both in A_1 or both in A_2. To keep the notation neutral, say they are both in A_2'. Since $A_2' \subseteq A_3$ or $A_3 \subseteq A_2'$, all of v_1, v_2, v_3 are in A_2' or they are all in A_3. Say they are all in A_3'. Continuing in this way, we arrive at one of the sets $A_1, \dots, A_n$, say A_n', such that all of $v_1, \dots, v_n$ are all in A_n'. The members of A_n' are linearly independent by assumption, and we obtain the contradiction $c_1 = \cdots = c_n = 0$. We conclude that A is linearly independent. Thus the chain $\mathcal{T}$ has an upper bound in S. By Zorn's Lemma, S has a maximal element, say M. By Proposition 2.1a, M is a basis of V containing E.

For (a), let E be the given spanning set, and let $\mathcal{S}$ be the collection of all linearly independent subsets of V that are contained in E. Partially order $\mathcal{S}$ by inclusion upward. The set $\mathcal{S}$ is nonempty because $\varnothing$ is in $\mathcal{S}$. Let $\mathcal{T}$ be a chain in $\mathcal{S}$, and let A be the union of the members of $\mathcal{T}$. We show that A is in $\mathcal{S}$, and then A is certainly an upper bound of $\mathcal{T}$. Because of its definition, A is contained in E, and the same argument as in the previous paragraph shows that A is linearly independent. Thus the chain $\mathcal{T}$ has an upper bound in S. By Zorn's Lemma, S has a maximal element, say M. Proposition 2.1a is not applicable, but its proof is easily adjusted to apply here to show that M spans V and hence is a basis: Given v in V, we are to prove that v lies is the linear span of M. First suppose that v is in E. If v is in M, there is nothing to prove. Since $M \cup \{v\}$ is contained in E, the assumed maximality implies that $M \cup \{v\}$ is not linearly independent, and hence $cv + c_1v_1 + \cdots + c_nv_n = 0$ for some scalars $c, c_1, \dots, c_n$ not all 0 and for some vectors $v_1, \dots, v_n$ in M. The scalar c cannot be 0 since M is linearly independent. Thus $v = -c^{-1}c_1v_1 - \cdots - c^{-1}c_nv_n$, and v is exhibited as in the linear span of M. Consequently every member of E lies in the linear span of M. Now suppose that v is not in E. Since every member of V lies in the linear span of E, every member of V lies in the linear span of M.

Conclusion (c) follows from (a) by taking the spanning set to be V; alternatively it follows from (b) by taking the linearly independent set to be $\varnothing$.

For (d), let $A = \{v_\alpha\}$ and $B = \{w_\beta\}$ be two bases of V. Each member a of A can be written as $a = c_1w_{\beta_1} + \cdots + c_nw_{\beta_n}$ uniquely with the scalars $c_1, \dots, c_n$ nonzero and with each w_{β_j} in B. Let B_a be the finite subset $\{w_{\beta_1}, \dots, w_{\beta_n}\}$. Then we have associated to each member of A a finite subset B_a of B. Let us see that $\bigcup_{a\in A} B_a = B$. If b is in B, then the linear span of $B - \{b\}$ is not all of V. Thus some v in V is not in this span. Expand v in terms of A as $v = d_1v_{\alpha_1} + \cdots + d_mv_{\alpha_m}$ with all $d_j \neq 0$. Since v is not in the linear span of $B - \{b\}$, some $a_0 = v_{\alpha_{j_0}}$ with $1 \leq j_0 \leq m$ is not in this linear span. Then b is in B_{a_0}, and we conclude that $B = \bigcup_{a\in A} B_a$. By the corollary near the end of Section A6 of the appendix, $\operatorname{card} B \leq \operatorname{card} A$. Reversing the roles of A and B, we obtain $\operatorname{card} A \leq \operatorname{card} B$. By the Schroeder–Bernstein Theorem, A and B have the same cardinality. This proves (d). $\square$

Now let us go through the results of the chapter and see how many of them extend to the infinite-dimensional case and why. It is possible but not very useful in the infinite-dimensional case to associate an infinite "matrix" to a linear map when bases or ordered bases are specified for the domain and range. Because this association is not very useful, we shall not attempt to extend any of the results concerning matrices. The facts concerning extensions of results just dealing with dimensions and linear maps are as follows:

COROLLARY 2.5. If V is any vector space and U is a vector subspace, then $\dim U \leq \dim V$.

In fact, take a basis of U and extend it to a basis of V; a basis of U is then exhibited as a subset of a basis of V, and the conclusion about cardinal-number dimensions follows.

PROPOSITION 2.13. Let U and V be vector spaces over $\mathbb{F}$, and let Γ be a basis of U. Then to each function $\ell : \Gamma \to V$ corresponds one and only one linear map $L : U \to V$ such that $L\big|_\Gamma = \ell$.

In fact, the proof given in Section 3 is valid with no assumption about finite dimensionality.

COROLLARY 2.15. If $L : U \to V$ is a linear map between vector spaces over $\mathbb{F}$, then

$$\dim(\mathrm{domain}(L)) = \dim(\mathrm{kernel}(L)) + \dim(\mathrm{image}(L)).$$

In fact, this formula remains valid, but the earlier proof via matrices has to be replaced. Instead, take a basis $\{v_\alpha \mid \alpha \in A\}$ of the kernel and extend it to a basis $\{v_\alpha \mid \alpha \in S\}$ of the domain. It is routine to check that $\{L(v_\alpha) \mid \alpha \in S - A\}$ is a basis of the image of L.

THEOREM 2.16 (part). The composition of two linear maps is linear.

In fact, the proof in Section 3 remains valid with no assumption about finite dimensionality.

PROPOSITION 2.18. Two vector spaces over $\mathbb{F}$ are isomorphic if and only if they have the same cardinal-number dimension.

In fact, this result follows from Proposition 2.13 just as it did in the finite-dimensional case; the only changes that are needed in the argument in Section 3 are small adjustments of the notation. Of course, one must not overinterpret this result on the basis of the remark with Theorem 2.42: two vector spaces with dimension $+\infty$ need not be isomorphic. Despite the apparent definitive sound of Proposition 2.18, one must not attach too much significance to it; vector spaces that arise in practice tend to have some additional structure, and an isomorphism based merely on equality of dimensions need not preserve the additional structure.

PROPOSITION 2.19. If V is a vector space and V' is its dual, then $\dim V \leq \dim V'$. (In the infinite-dimensional case we do not have equality.)

In fact, take a basis $\{v_\alpha\}$ of V. If for each α we define $v'_\alpha(v_\beta) = \delta_{\alpha\beta}$ and use Proposition 2.13 to form the linear extension v'_α, then the set $\{v'_\alpha\}$ is a linearly independent subset of V' that is in one-one correspondence with the basis of V. Extending $\{v'_\alpha\}$ to a basis of V', we obtain the result.

PROPOSITION 2.20. Let V be a vector space, and let U be a vector subspace of V. Then

(b) every linear functional on U extends to a linear functional on V,
(c) whenever v_0 is a member of V that is not in U, there exists a linear functional on V that is 0 on U and is 1 on v_0.

Conclusion (a) of the original Proposition 2.20, which concerns annihilators, does not extend to the infinite-dimensional case.

To prove (b) without the finite dimensionality, let u' be a given linear functional on U, let $\{u_\alpha\}$ be a basis of U, and let $\{v_\beta\}$ be a subset of V such that $\{u_\alpha\} \cup \{v_\beta\}$ is a basis of V. Define $v'(u_\alpha) = u'(u_\alpha)$ for each α and $v'(v_\beta) = 0$ for each β. Using Proposition 2.13, let v' be the linear extension to a linear functional on V. Then v' has the required properties.

To prove (c) without the finite dimensionality, we take a basis $\{u_\alpha\}$ of U and extend $\{u_\alpha\} \cup \{v_0\}$ to a basis of V. Define v' to equal 0 on each u_α, to equal 1 on v_0, and to equal 0 on the remaining members of the basis of V. Then the linear extension of v' to V is the required linear functional.

PROPOSITION 2.22. If V is any vector space over $\mathbb{F}$, then the canonical map $\iota : V \to V''$ is one-one. The canonical map is not onto V'' if V is infinite-dimensional.

The proof that it is one-one given in Section 4 is applicable in the infinite-dimensional case since we know from Theorem 2.42 that any linearly independent subset of V can be extended to a basis. For the second conclusion when V has a countably infinite basis, see Problem 31 at the end of the chapter.

PROPOSITION 2.23 THROUGH COROLLARY 2.29. For these results about quotients, the only place that finite dimensionality played a role was in the dimension formulas, Corollaries 2.24 and 2.29. We restate these two results separately.

COROLLARY 2.24. If V is a vector space over $\mathbb{F}$ and U is a vector subspace, then

(a) $\dim V = \dim U + \dim(V/U)$,
(b) the subspace U is the kernel of some linear map defined on V.

The proof in Section 5 requires no changes: Let q be the quotient map. The linear map q meets the conditions of (b). For (a), take a basis of U and extend to a basis of V. Then the images under q of the additional vectors form a basis of V/U.

COROLLARY 2.29. Let M and N be vector subspaces of a vector space V over $\mathbb{F}$. Then

$$\dim(M+N)+\dim(M\cap N)=\dim M+\dim N.$$

In fact, Corollary 2.24a gives us $\dim(M+N)=\dim((M+N)/M)+\dim M$. Substituting $\dim((M+N)/M)=\dim(N/(M\cap N))$ from Theorem 2.28 and adding $\dim(M\cap N)$ to both sides, we obtain $\dim(M+N)+\dim(M\cap N)=\dim(M\cap N)+\dim(N/(M\cap N))+\dim M$. The first two terms on the right side add to $\dim N$ by Corollary 2.24a, and the result follows.

PROPOSITIONS 2.30 THROUGH 2.33. These results about direct products and direct sums did not assume any finite dimensionality.

The determinants of Sections 7–8 have no infinite-dimensional generalization, and Proposition 2.41 is the only result in those two sections with a valid infinite-dimensional analog. The valid analog in the infinite-dimensional case is that eigenvectors for distinct eigenvalues under a linear map are linearly independent. The proof given for Proposition 2.41 in Section 8 adapts to handle this analog, provided we interpret components $v_i^{(j)}$ of a vector v_i as the coefficients needed to expand v_i in a basis of the underlying vector space.

10. Problems

1. Determine bases of the following subsets of $\mathbb{R}^3$:
   (a) the plane $3x-2y+5z=0$,
   (b) the line $\left\{\begin{array}{l} x=2t \\ y=-t \\ z=4t \end{array}\right\}$, where $-\infty<t<\infty$.
2. This problem shows that the associativity law in the definition of "vector space" implies certain more complicated formulas of which the stated law is a special case. Let $v_1,\dots,v_n$ be vectors in a vector space V. The only vector-space properties that are to be used in this problem are associativity of addition and the existence of the 0 element.
   (a) Define $v_{(k)}$ inductively upward by $v_{(0)}=0$ and $v_{(k)}=v_{(k-1)}+v_k$, and define $v^{(l)}$ inductively downward by $v^{(n+1)}=0$ and $v^{(l)}=v_l+v^{(l+1)}$. Prove that $v_{(k)}+v^{(k+1)}$ is always the same element for $0\le k\le n$.
   (b) Prove that the same element of V results from any way of inserting parentheses in the sum $v_1+\cdots+v_n$ so that each step requires the addition of only two members of V.

3. This problem shows that the commutative and associative laws in the definition of "vector space" together imply certain more complicated formulas of which the stated commutative law is a special case. Let $v_1, \dots, v_n$ be vectors in a vector space V. The only vector-space properties that are to be used in this problem are commutativity of addition and the properties in the previous problem. Because of the previous problem, $v_1 + \cdots + v_n$ is a well-defined element of V, and it is not necessary to insert any parentheses in it. Prove that $v_1 + v_2 + \cdots + v_n = v_{\sigma(1)} + v_{\sigma(2)} + \cdots + v_{\sigma(n)}$ for each permutation σ of $\{1, \dots, n\}$.

4. For the matrix $A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 6 \\ 0 & 0 & -8 \end{pmatrix}$, find
   (a) a basis for the row space,
   (b) a basis for the column space, and
   (c) the rank of the matrix.

5. Let A be an n-by-n matrix of rank one. Prove that there exists an n-dimensional column vector c and an n-dimensional row vector r such that $A = cr$.

6. Let A be a k-by-n matrix, and let R be a reduced row-echelon form of A.
   (a) Prove for each r that the rows of R whose first r entries are 0 form a basis for the vector subspace of all members of the row space of A whose first r entries are 0.
   (b) Prove that the reduced row-echelon form of A is unique in the sense that any two sequences of steps of row reduction lead to the same reduced form.

7. Let E be an finite set of N points, let V be the N-dimensional vector space of all real-valued functions on E, and let n be an integer with $0 < n \le N$. Suppose that U is an n-dimensional subspace of V. Prove that there exists a subset D of n points in E such that the vector space of restrictions to D of the members of U has dimension n.

8. A linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ is given in the standard ordered basis by the matrix $\begin{pmatrix} -6 & -12 \\ 6 & 11 \end{pmatrix}$. Find the matrix of L in the ordered basis $\left\{ \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \begin{pmatrix} -4 \\ 3 \end{pmatrix} \right\}$.

9. Let V be the real vector space of all polynomials in x of degree ≤ 2, and let $L : V \to V$ be the linear map $I - D^2$, where I is the identity and D is the differentiation operator d/dx. Prove that L is invertible.

10. Let A be in $M_{km}(\mathbb{C})$ and B be in $M_{mn}(\mathbb{C})$. Prove that

$$\operatorname{rank}(AB) \le \max(\operatorname{rank} A, \operatorname{rank} B).$$

11. Let A be in $M_{kn}(\mathbb{C})$ with $k > n$. Prove that there exists no B in $M_{nk}(\mathbb{C})$ with $AB = I$.

12. Let A be in $M_{kn}(\mathbb{C})$ and B be in $M_{nk}(\mathbb{C})$. Give an example with $k = n$ to show that $\operatorname{rank}(AB)$ need not equal $\operatorname{rank}(BA)$.

13. With the differential equation $y''(t) = y(t)$ in Example 2 of Section 3, two examples of linear functionals on the vector space of solutions are given by $\ell_1(y) = y(0)$ and $\ell_2(y) = y'(0)$. Find a basis of the space of solutions such that $\{\ell_1, \ell_2\}$ is the dual basis.

14. Suppose that a vector space V has a countably infinite basis. Prove that the dual V' has an uncountable linearly independent set.

15. (a) Give an example of a vector space and three vector subspaces L, M, and N such that $L \cap (M + N) \neq (L \cap M) + (L \cap N)$.
    (b) Show that inclusion always holds in one direction in (a).
    (c) Show that equality always holds in (a) if $L \supseteq M$.

16. Construct three vector subspaces M, N_1, and N_2 of a vector space V such that $M \oplus N_1 = M \oplus N_2 = V$ but $N_1 \neq N_2$. What is the geometric picture corresponding to this situation?

17. Suppose that x, y, u, and v are vectors in $\mathbb{R}^4$; let M and N be the vector subspaces of $\mathbb{R}^4$ spanned by $\{x, y\}$ and $\{u, v\}$, respectively. In which of the following cases is it true that $\mathbb{R}^4 = M \oplus N$?
    (a) $x = (1, 1, 0, 0)$, $y = (1, 0, 1, 0)$, $u = (0, 1, 0, 1)$, $v = (0, 0, 1, 1)$;
    (b) $x = (-1, 1, 1, 0)$, $y = (0, 1, -1, 1)$, $u = (1, 0, 0, 0)$, $v = (0, 0, 0, 1)$;
    (c) $x = (1, 0, 0, 1)$, $y = (0, 1, 1, 0)$, $u = (1, 0, 1, 0)$, $v = (0, 1, 0, 1)$.

18. Section 6 gave definitions and properties of projections and injections associated with the direct sum of two vector spaces. Write down corresponding definitions and properties for projections and injections in the case of the direct sum of n vector spaces, n being an integer > 2.

19. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map with $\ker T \cap \operatorname{image} T = 0$.
    (a) Prove that $\mathbb{R}^n = \ker T \oplus \operatorname{image} T$.
    (b) Prove that the condition $\ker T \cap \operatorname{image} T = 0$ is satisfied if $T^2 = T$.

20. If V_1 and V_2 are two vector spaces over $\mathbb{F}$, prove that $(V_1 \oplus V_2)'$ is canonically isomorphic to $V_1' \oplus V_2'$.

21. Suppose that M is a vector subspace of a vector space V and that $q : V \to V/M$ is the quotient map. Corresponding to each linear functional y on V/M is a linear functional z on V given by $z = yq$. Why is the correspondence $y \mapsto z$ an isomorphism between $(V/M)'$ and $\operatorname{Ann} M$?

22. Let M be a vector subspace of the vector space V, and let $q : V \to V/M$ be the quotient map. Suppose that N is a vector subspace of V. Prove that $V = M \oplus N$ if and only if the restriction of q to N is an isomorphism of N onto V/M.

23. For a square matrix A of integers, prove that the inverse has integer entries if and only if $\det A = \pm 1$.

24. Let A be in $M_{kn}(\mathbb{C})$, and let $r = \operatorname{rank} A$. Prove that r is the largest integer such that there exist r row indices $i_1, \dots, i_r$ and r column indices $j_1, \dots, j_r$ for which the r-by-r matrix formed from these rows and columns of A has nonzero determinant. (Educational note: This problem characterizes the subset of matrices of rank $\leq r - 1$ as the set in which all determinants of r-by-r submatrices are zero.)

25. Suppose that a linear combination of functions $t \mapsto e^{ct}$ with c real vanishes for every integer $t \geq 0$. Prove that it vanishes for every real t.

26. Find all eigenvalues and eigenvectors of $A = \begin{pmatrix} 0 & 1 \\ -6 & 5 \end{pmatrix}$.

27. Let A and C be n-by-n matrices with C invertible. By making a direct calculation with the entries, prove that $\operatorname{Tr}(C^{-1}AC) = \operatorname{Tr} A$.

28. Find the characteristic polynomial of the n-by-n matrix $\begin{pmatrix} 0 & 1 & 0 & 0 & & 0 & 0 \\ 0 & 0 & 1 & 0 & & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 0 \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ a_0 & a_1 & a_2 & a_3 & \cdots & a_{n-2} & a_{n-1} \end{pmatrix}$.

29. Let A and B be in $M_{nn}(\mathbb{C})$.
   (a) Prove under the assumption that A is invertible that $\det(\lambda I - AB) = \det(\lambda I - BA)$.
   (b) By working with $A + \epsilon I$ and letting ϵ tend to 0, show that the assumption in (a) that A is invertible can be dropped.

30. In proving Theorem 2.42a, it is tempting to argue by considering all spanning subsets of the given set, ordering them by inclusion downward, and seeking a minimal element by Zorn's Lemma. Give an example of a chain in this ordering that has no lower bound, thereby showing that this line of argument cannot work.

Problems 31–34 concern annihilators. Let V be a vector space, let M and N be vector subspaces, and let $\iota : V \to V''$ be the canonical map.

31. If V has an infinite basis, how can we conclude that ι does not carry V onto V''?

32. Prove that $\operatorname{Ann}(M + N) = \operatorname{Ann} M \cap \operatorname{Ann} N$.

33. Prove that $\operatorname{Ann}(M \cap N) = \operatorname{Ann} M + \operatorname{Ann} N$.

34. (a) Prove that $\iota(M) \subseteq \operatorname{Ann}(\operatorname{Ann} M)$.
   (b) Prove that equality holds in (a) if V is finite-dimensional.
   (c) Give an infinite-dimensional example in which equality fails in (a).

Problems 35–39 concern operations by blocks within matrices.

35. Let A be a k-by-m matrix of the form $A = \begin{pmatrix} A_1 & A_2 \end{pmatrix}$, where A_1 has size k-by-m_1, A_2 has size k-by-m_2, and $m_1 + m_2 = m$. Let B by an m'-by-n matrix of the form $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$, where B_1 has size m_1'-by-n, B_2 has size m_2'-by-n, and $m_1' + m_2' = m'$.
    (a) If $m_1 = m_1'$ and $m_2 = m_2'$, prove that $AB = A_1B_1 + A_2B_2$.
    (b) If $k = n$, prove that $BA = \begin{pmatrix} B_1A_1 & B_1A_2 \\ B_2A_2 & B_2A_2 \end{pmatrix}$.
    (c) Deduce a general rule for block multiplication of matrices that are in 2-by-2 block form.

36. Let A be in $M_{kk}(\mathbb{C})$, B be in $M_{kn}(\mathbb{C})$, and D be in $M_{nn}(\mathbb{C})$. Prove that $\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det A \det D$.

37. Let A, B, C, and D be in $M_{nn}(\mathbb{C})$. Suppose that A is invertible and that $AC = CA$. Prove that $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD - CB)$.

38. Let A be in $M_{kn}(\mathbb{C})$ and B be in $M_{nk}(\mathbb{C})$ with $k \le n$. Let I_k be the k-by-k identity, and let I_n be the n-by-n identity. Using Problem 29, prove that $\det(\lambda I_n - BA) = \lambda^{n-k} \det(\lambda I_k - AB)$.

39. Prove the following block-form generalization of the expansion-in-cofactors formula. For each subset S of $\{1, \dots, n\}$, let S^c be the complementary subset within $\{1, \dots, n\}$, and let $\operatorname{sgn}(S, S^c)$ be the sign of the permutation that carries $(1, \dots, n)$ to the members of S in order, followed by the members of S^c in order. Fix k with $1 \le k \le n-1$, and let the subset S have $|S| = k$. For an n-by-n matrix A, define $A(S)$ to be the square matrix of size k obtained by using the rows of A indexed by $1, \dots, k$ and the columns indexed by the members of S. Let $\widehat{A}(S)$ be the square matrix of size $k-1$ obtained by using the rows of A indexed by $k+1, \dots, n$ and the columns indexed by the members of S^c. Prove that
$$\det A = \sum_{\substack{S \subseteq \{1,\dots,n\}, \\ |S| = k}} \operatorname{sgn}(S, S^c) \det A(S) \det \widehat{A}(S).$$

Problems 40–44 compute the determinants of certain matrices known as **Cartan matrices**. These have geometric significance in the theory of Lie groups.

40. Let A_n be the n-by-n matrix $\begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 2 & \cdots & 0 & 0 \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}$. Using expansion in cofactors about the last row, prove that $\det A_n = 2\det A_{n-1} - \det A_{n-2}$ for $n \geq 3$.

41. Computing $\det A_1$ and $\det A_2$ directly and using the recursion in Problem 40, prove that $\det A_n = n + 1$ for $n \geq 1$.
42. Let C_n for $n \geq 2$ be the matrix A_n except that the $(1, 2)^{\text{th}}$ entry is changed from -1 to -2.
    (a) Expanding in cofactors about the last row, prove that the argument of Problem 40 is still applicable when $n \geq 4$ and a recursion formula for $\det C_n$ results with the same coefficients.
    (b) Computing $\det C_2$ and $\det C_3$ directly and using the recursion equation in (a), prove that $\det C_n = 2$ for $n \geq 2$.
43. Let D_n for $n \geq 3$ be the matrix A_n except that the upper left 3-by-3 piece is changed from $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$ to $\begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$.
    (a) Expanding in cofactors about the last row, prove that the argument of Problem 40 is still applicable when $n \geq 5$ and a recursion formula for $\det D_n$ results with the same coefficients.
    (b) Show that D_3 can be transformed into A_3 by suitable interchanges of rows and interchanges of columns, and conclude that $\det D_3 = \det A_3 = 4$.
    (c) Computing $\det D_4$ directly and using (b) and the recursion equation in (a), prove that $\det D_n = 4$ for $n \geq 3$.
44. Let E_n for $n \geq 4$ be the matrix A_n except that the upper left 4-by-4 piece is changed from $\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$ to $\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}$.
    (a) Expanding in cofactors about the last row, prove that the argument of Problem 40 is still applicable when $n \geq 6$ and a recursion formula for $\det E_n$ results with the same coefficients.
    (b) Show that E_4 can be transformed into A_4 by suitable interchanges of rows and interchanges of columns, and conclude that $\det E_4 = \det A_4 = 5$.
    (c) Show that E_5 can be transformed into D_5 by suitable interchanges of rows and interchanges of columns, and conclude that $\det E_5 = \det D_5 = 4$.
    (d) Using (b) and (c) and the recursion equation in (a), prove that $\det E_n = 9 - n$ for $n \geq 4$.

Problems 45–48 relate determinants to areas and volumes. They begin by showing how a computation of an area in $\mathbb{R}^2$ leads to a determinant, they then show how knowledge of the answer and of the method of row reduction illuminate the result, and finally they indicate how the result extends to $\mathbb{R}^3$. If u and v are vectors in $\mathbb{R}^2$, let us say that the parallelogram determined by u and v is the parallelogram with vertices 0, u, v, and $u+v$. If u, v, and w are in $\mathbb{R}^3$, the parallelepiped determined by u, v, and w is the parallelepiped with vertices 0, u, v, w, $u+v$, $u+w$, $v+w$, and $u+v+w$.

45. The area of a trapezoid is the product of the average of the two parallel sides by the distance between the parallel sides. Compute the area of the parallelogram determined by $u = \begin{pmatrix} a \\ c \end{pmatrix}$ and $v = \begin{pmatrix} b \\ d \end{pmatrix}$ in the diagram below as the area of a large rectangle minus the area of two trapezoids minus the area of two triangles, recognizing the answer as $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ except for a minus sign. To what extent is the answer dependent on the picture?

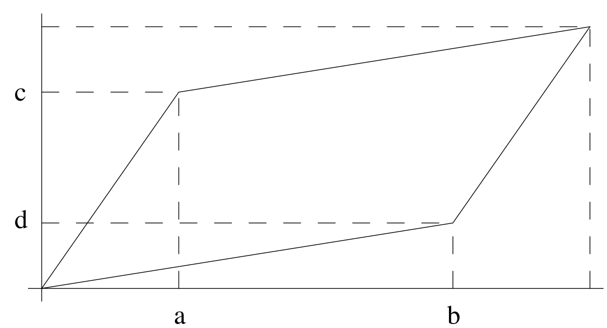


FIGURE 2.6. Area of a parallelogram as a difference of areas.

46. What is the geometric effect on the parallelogram of replacing the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, i.e., of right-multiplying $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by $\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$? What does this change do to the area? What algebraic operation does this change correspond to?

47. Answer the same questions as in Problem 46 for right multiplication by the matrices $\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} q & 0 \\ 0 & 1 \end{pmatrix}$ for a nonzero number q, and $\begin{pmatrix} 1 & 0 \\ 0 & r \end{pmatrix}$ for a nonzero number r.

48. Explain on the basis of Problems 45–47 why if three column vectors u, v, and w in $\mathbb{R}^3$ are assembled into a 3-by-3 matrix A and A is invertible, then the volume of the parallelepiped determined by u, v, and w has to be $|\det A|$.