Linear Algebra Notes

Wuwei Yuan

December 25, 2020

Contents

1 Vector Spaces

1.1 Introduction

Parallelogram Law for Vector Addition. *The sum of two vectors x and y that act at the same point P is the vector beginning at P that is represented by the diagonal of parallelogram having x and y as adjacent sides.*

1.2 Vector Spaces

Definition (Field). *Formally, a field is a set F together with two binary operations on F called addition and multiplication. A binary operation on F is a mapping $F \times F \to F$, that is, a correspondence that associates with each ordered pair of elements of F a uniquely determined element of F. The result of the addition of a and b is called the sum of a and b, and is denoted $a+b$. Similarly, the result of the multiplication of a and b is called the product of a and b, and is denoted ab or $a \cdot b$. These operations are required to satisfy the following properties, referred to as field axioms. In these axioms, a, b, and c are arbitrary elements of the field F.*

- *Associativity of addition and multiplication: $a + (b + c) = (a + b) + c$, and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.*

- *Commutativity of addition and multiplication: $a+b = b+a$, and $a \cdot b = b \cdot a$.*

- *Additive and multiplicative identity: there exist two different elements 0 and 1 in F such that $a + 0 = a$ and $a \cdot 1 = a$.*

- *Additive inverses: for every a in F, there exists an element in F, denoted $-a$, called the additive inverse of a, such that $a + (-a) = 0$.*

- *Multiplicative inverses: for every $a \neq 0$ in F, there exists an element in F, denoted by a^{-1} or $1/a$, called the multiplicative inverse of a, such that $a \cdot a^{-1} = 1$.*

- *Distributivity of multiplication over addition: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.*

Definitions. *A vector space (or linear space) V over a field F consists of a set on which two operations (called addition and scalar multiplication, respectively) are defined so that for each pair of elements x, y, in V there is a unique element $x + y$ in V, and for each element a in F and each element x in V there is a unique element ax in V, such that the following conditions hold.*

(VS 1) *For all x, y in V, $x + y = y + x$ (commutativity of addition).*

(VS 2) *For all x, y, z in V, $(x + y) + z = x + (y + z)$ (associativity of addition).*

(VS 3) *There exists an element in V denoted by 0 such that $x+0 = x$ for each x in V.*

(VS 4) For each element x in V there exists an element y in V such that $x + y = 0$

(VS 5) For each element x in V, $1x = x$.

(VS 6) For each pair of elements a, b in F and each element x in V, $(ab)x = a(bx)$.

(VS 7) For each element a in F and each pair of elements x, y in V, $a(x + y) = ax + ay$.

(VS 8) For each pair of elements a, b in F and each element x in V, $(a + b)x = ax + bx$.

The elements $x + y$ and ax are called the sum of x and y and the product of a and x, respectively.

The elements of the field F are called scalars and the elements of the vector space V are called vectors.

Theorem 1.1 (Cancellation Law for Vector Addition)**.** *If x, y, and z are vectors in a vector space V such that $x + z = y + z$, then $x = y$.*

Corollary 1. *The vector 0 described in (VS 3) is unique.*

Corollary 2. *The vector y described in (VS 4) is unique.*

The vector 0 in (VS 3) is called the zero vector of V, and the vector y in (VS 4) (that is, the unique vector such that x+y = 0) is called the additive inverse of x and is denoted by −x.

Theorem 1.2. *In any vector space V, the following statements are true:*

(a) $0x = 0$ for each $x \in V$.

(b) $(-a)x = -(ax) = a(-x)$ for each $a \in F$ and each $x \in V$.

(c) $a0 = 0$ for each $a \in F$.

1.3 Subspaces

Definition. *A subset W of a vector space V over a field F is called a subspace of V if W is a vector space over F with the operations of addition and scalar multiplication defined on V.*

In any vector space V, note that V and {0} are subspaces. The latter is called the zero subspace of V.

Theorem 1.3. *Let V be a vector space and W a subset of V. Then W is a subspace of V if and only if the following three conditions hold for the operations defined in V.*

(a) $0 \in W$.

(b) $x + y \in W$ *whenever* $x \in W$ *and* $y \in W$.

(c) $cx \in W$ *whenever* $c \in F$ *and* $x \in W$.

The **transpose** A^t of an $m \times n$ matrix A is the $n \times m$ matrix obtained from A by interchanging the rows with the columns; that is, $(A^t)_{ij} = A_{ji}$.

Theorem 1.4. *Any intersection of subspaces of a vector space V is a subspace of V.*

Definition. *If S_1 and S_2 are nonempty subsets of a vector space V, then the sum of S_1 and S_2, denoted $S_1 + S_2$, is the set $\{x + y : x \in S1 \text{ and } y \in S2\}$.*

Definition. *A vector space V is called the direct sum of W_1 and W_2 if W_1 and W_2 are subspaces of V such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$. We denote that V is the direct sum of W_1 and W_2 by writing $V = W_1 \oplus W_2$.*

Exercise. *$W_1 + W_2$ is a subspace of V that contains both W_1 and W_2.*

Exercise. *Any subspace of V that contains both W_1 and W_2 must also contain $W_1 + W_2$.*

Definitions. *Let W be a subspace of a vector space V over a field F. For any $v \in V$ the set $\{v\} + W = \{v + w : w \in W\}$ is called the **coset** of W **containing** v. It is customary to denote this coset by $v + W$ rather than $\{v\} + W$.*

Addition and scalar multiplication by scalars of F can be defined in the collection $S = \{v + W : v \in V\}$ of all cosets of W as follows:

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W \tag{1}$$

for all $v_1, v_2 \in V$ and

$$a(v + W) = av + W \tag{2}$$

for all $v \in V$ and $a \in F$.

*The set S is a vector space with the preceding operations. This vector space is called the **quotient space of** V **modulo** W and is denoted by V/W.*

1.4   Linear Combinations and Systems of Linear Equations

Definition. *Let V be a vector space and S a nonempty subset of V. A vector $v \in V$ is called a linear combination of vectors of S if there exist a finite number of vectors $u_1, u_2, \ldots, u_n$ in S and scalars $a_1, a_2, \ldots, a_n$ in F such that $v = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n$. In this case we also say that v is a linear combination of $u_1, u_2, \ldots, u_n$ and call $a_1, a_2, \ldots, a_n$ the coefficients of the linear combination.*

Definition. *Let S be a nonempty subset of a vector space V. The span of S, denoted span(S), is the set consisting of all linear combinations of the vectors in S. For convenience, we define $span(\varnothing) = \{0\}$.*

Theorem 1.5. *The span of any subset S of a vector space V is a subspace of V. Moreover, any subspace of V that contains S must also contain the span of S.*

Definition. *A subset S of a vector space V generates (or spans) V if span(S) = V. In this case, we also say that the vectors of S generate (or span) V.*

1.5   Linear Dependence and Linear Independence

Definition. *A subset S of a vector space V is called linearly dependent if there exist a finite number of distinct vectors $u_1, u_2, \ldots, u_n$ in S and scalars $a_1, a_2, \ldots, a_n$, not all zero, such that*

$$a_1 u_1 + a_2 u_2 + \cdots + a_n u_n = 0.$$

In this case we also say that the vectors of S are linearly dependent.

For any vectors $u_1, u_2, \ldots, u_n$, we have $a_1 u_1 + a_2 u_2 + \cdots + a_n u_n = 0$ if $a_1 = a_2 = \cdots = a_n = 0$. We call this the trivial representation of 0 as a linear combination of $u_1, u_2, \ldots, u_n$.

Definition. *A subset S of a vector space that is not linearly dependent is called linearly independent. As before, we also say that the vectors of S are linearly independent.*

The following facts about linearly independent sets are true in any vector space.

1. The empty set is linearly independent, for linearly dependent sets must be nonempty.

2. A set consisting of a single nonzero vector is linearly independent. For if u is linearly dependent, then au = 0 for some nonzero scalar a. Thus

$$u = a^{-1}(au) = a^{-1}0 = 0.$$

3. A set is linearly independent if and only if the only representations of 0 as linear combinations of its vectors are trivial representations.

Theorem 1.6. *Let V be a vector space, and let $S_1 \subseteq S_2 \subseteq V$. If S_1 is linearly dependent, then S_2 is linearly dependent.*

Corollary. *Let V be a vector space, and let $S_1 \subseteq S_2 \subseteq V$. If S_2 is linearly independent, then S_1 is linearly independent.*

Theorem 1.7. *Let S be a linearly independent subset of a vector space V, and let v be a vector in V that is not in S. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in span(S)$.*

1.6 Base and Dimension

Definition. *A **basis** β for a vector space V is a linearly independent subset of V that generates V. If β is a basis for V, we also say that the vectors of β form a basis for V.*

Theorem 1.8. *Let V be a vector space and $\beta = \{u_1, u_2, \ldots, u_n\}$ be a subset of V. Then β is a basis for V if and only if each $v \in V$ can be uniquely expressed as a linear combination of vectors of β, that is, can be expressed in the form*

$$v = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n$$

for unique scalars $a_1, a_2, \ldots, a_n$.

Theorem 1.9. *If a vector space V is generated by a finite set S, then some subset of S is a basis for V. Hence V has a finite basis.*

Theorem 1.10 (Replacement Theorem)**.** *Let V be a vector space that is generated by a set G containing exactly n vectors, and let L be a linearly independent subset of V containing exactly m vectors. Then $m \leq n$ and there exists a subset H of G containing exactly $n - m$ vectors such that $L \cup H$ generates V.*

Corollary 1. *Let V be a vector space having a finite basis. Then every basis for V contains the same number of vectors.*

Definitions. *A vector space is called **finite-dimensional** if it has a basis consisting of a finite number of vectors. The unique number of vectors in each basis for V is called the **dimension** of V and is denoted by dim(V). A vector space that is not finite-dimensional is called **infinite-dimensional**.*

Corollary 2. *Let V be a vector space with dimension n.*

  (a) *Any finite generating set for V contains at least n vectors, and a generating set for V that contains exactly n vectors is a basis for V.*

  (b) *Any linearly independent subset of V that contains exactly n vectors is a basis for V.*

  (c) *Every linearly independent subset of V can be extended to a basis for V.*

Theorem 1.11. *Let W be a subspace of a finite-dimensional vector space V. Then W is finite-dimensional and $dim(W) \leq dim(V)$. Moreover, if $dim(W) = dim(V)$, then $V = W$.*

Corollary. *If W is a subspace of a finite-dimensional vector space V, then any basis for W can be extended to a basis for V.*

1.7 Maximal Linearly Independent Subsets*

Definition. *Let F be a family of sets. A member M of F is called maximal (with respect to set inclusion) if M is contained in no member of F other than M itself.*

Definition. *A collection of sets C is called a chain (or nest or tower) if for each pair of sets A and B in C, either A ⊆ B or B ⊆ A.*

Maximal Principle. *Let F be a family of sets. If, for each chain C ⊆ F, there exists a member of F that contains each member of C, then F contains a maximal member.*

Definition. *Let S be a subset of a vector space V. A maximal linearly independent subset of S is a subset B of S satisfying both of the following conditions.*

  (a) B is linearly independent.

  (b) The only linearly independent subset of S that contains B is B itself

A basis β for a vector space V is a maximal linearly independent subset of V, because 1. β is linearly independent by definition. 2. If v ∈ V and v ∉ β, then β ∪ v is linearly dependent by Theorem 1.7 (p. 39) because span(β) = V.

Theorem 1.12. *Let V be a vector space and S a subset that generates V. If β is a maximal linearly independent subset of S, then β is a basis for V.*

Theorem 1.13. *Let S be a linearly independent subset of a vector space V. There exists a maximal linearly independent subset of V that contains S.*

Corollary. *Every vector space has a basis.*

It can be shown, analogously to Corollary 1 of the replacement theorem (p. 46), that every basis for an infinite-dimensional vector space has the same cardinality.

Also there are many important exercise. See textbook P62.

2 Linear Transformations and Matrices

2.1 Linear Transformations, Null Spaces, and Ranges

Definition. *Let V and W be vector spaces (over F). We call a function T: V → W a linear transformation from V to W if, for all $x, y \in V$ and $c \in F$, we have*
  (a) T(x + y) = T(x) + T(y) and
  (b) T(cx) = cT(x).

We often simply call T **linear**. The reader should verify the following properties of a function $T : V \to W$. (See Exercise 7.)

  1. If T is linear, then $T(0) = 0$.

2. T is linear if and only if $T(cx + y) = cT(x) + T(y)$ for all $x, y \in V$ and $c \in F$.

3. If T is linear, then $T(x-y) = T(x)-T(y)$ for all $x, y \in V$.

4. T is linear if and only if, for $x_1, x_2, ..., x_n \in V$ and $a_1, a_2, \ldots, a_n \in F$, we have

$$T\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i T(x_i)$$

For vector spaces V and W (over F), we define the **identity transformation** $I_V : V \to V$ by $I_V(x) = x$ for all $x \in V$ and the **zero transformation** $T_0 : V \to W$ by $T_0(x) = 0$ for all $x \in V$. It is clear that both of these transformations are linear. We often write I instead of I_V.

Definitions. *Let V and W be vector spaces, and let T: $V \to W$ be linear. We define the **null space** (or **kernel**) N(T) of T to be the set of all vectors x in V such that T(x) = 0 ; that is, $N(T) = \{x \in V : T(x) = 0\}$.*
*We define the **range** (or **image**) R(T) of T to be the subset of W consisting of all images (under T) of vectors in V; that is, $R(T) = \{T(x) : x \in V\}$.*

Theorem 2.1. *Let V and W be vector spaces and T: $V \to W$ be linear. Then N(T) and R(T) are subspaces of V and W, respectively.*

Theorem 2.2. *Let V and W be vector spaces, and let T: $V \to W$ be linear. If $\beta = \{v_1, v_2, \ldots, v_n\}$ is a basis for V, then $R(T) = span(T(\beta)) = span(\{T(v_1), T(v_2), \ldots, T(v_n)\})$*

It should be noted that Theorem 2.2 is true if β is infinite, that is, $R(T) = span(\{T(v) : v \in \beta\})$.

Definitions. *Let V and W be vector spaces, and let $T : V \to W$ be linear. If N(T) and R(T) are finite-dimensional, then we define the **nullity** of T, denoted nullity(T), and the **rank** of T, denoted rank(T), to be the dimensions of N(T) and R(T), respectively.*

Theorem 2.3 (Dimension Theorem)**.** *Let V and W be vector spaces, and let T: $V \to W$ be linear. If V is finite-dimensional, then nullity(T) + rank(T) = dim(V).*

Theorem 2.4. *Let V and W be vector spaces, and let T: $V \to W$ be linear. Then T is one-to-one if and only if $N(T) = \{0\}$.*

Theorem 2.5. *Let V and W be vector spaces of equal (finite) dimension, and let T: $V \to W$ be linear. Then the following are equivalent.*
(a) T is one-to-one.
(b) T is onto.
(c) rank(T) = dim(V).

9

Theorem 2.6. *Let V and W be vector spaces over F, and suppose that $\{v_1, v_2, \ldots, v_n\}$ is a basis for V. For $w_1, w_2, \ldots, w_n$ in W, there exists exactly one linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for $i = 1, 2, \ldots, n$.*

Corollary. *Let V and W be vector spaces, and suppose that V has a finite basis $\{v_1, v_2, \ldots, v_n\}$. If U,T: $V \to W$ are linear and $U(v_i) = T(v_i)$ for i = 1, 2,...,n, then U = T.*

Definition. *Let V be a vector space and W_1 and W_2 be subspaces of V such that $V = W_1 \oplus W_2$. (Recall the definition of direct sum given in the exercises of Section 1.3.) A function $T : V \to V$ is called the **projection on** W_1 **along** W_2 if, for $x = x_1 + x_2$ with $x_1 \in W_1$ and $x_2 \in W_2$, we have $T(x) = x_1$.*

Definition. *Let V be a vector space, and let T: $V \to V$ be linear. A subspace W of V is said to be T-**invariant** if $T(x) \in W$ for every $x \in W$, that is, $T(W) \in W$. If W is T-invariant, we define the **restriction of** T **on** W to be the function $T_W : W \to W$ defined by $T_W(x) = T(x)$ for all $x \in W$.*

2.2 The Matrix Representation of a Linear Transformation

Definition. *Let V be a finite-dimensional vector space. An **ordered basis** for V is a basis for V endowed with a specific order; that is, an ordered basis for V is a finite sequence of linearly independent vectors in V that generates V.*

For the vector space F^n, we call $\{e_1, e_2, \ldots, e_n\}$ the **standard ordered basis** for F^n. Similarly, for the vector space $P_n(F)$, we call $\{1, x, \ldots, x^n\}$ the **standard ordered basis** for $P_n(F)$.

Definition. *Let $ = \{u_1, u_2, \ldots, u_n\}$ be an ordered basis for a finite dimensional vector space V. For $x \in V$, let $a_1, a_2, \ldots, a_n$ be the unique scalars such that*

$$x = \sum_{i=1}^{n} a_i u_i$$

*We define the **coordinate vector of** x **relative to** β, denoted $[x]_\beta$, by*

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

The correspondence $x \to [x]_\beta$ provides us with a linear transformation from V to F^n.

Definition. *Suppose that V and W are finite-dimensional vector spaces with ordered bases $\beta = \{v_1, v_2, \cdots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_m\}$, respectively. Let*

$T: V \to W$ be linear. Then for each j, $1 \leq j \leq n$, there exist unique scalars $a_{ij} \in F$, $1 \leq i \leq m$, such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \ for \ 1 \leq j \leq n.$$

Using the notation above, we call the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$ the **matrix representation of** T **in the ordered bases** β and γ and write $A = [T]_\beta^\gamma$. If $V = W$ and $\beta = \gamma$, then we write $A = [T]_\beta$.

Definition. Let $T, U : V \to W$ be arbitrary functions, where V and W are vector spaces over F, and let $a \in F$. We define $T + U : V \to W$ by $(T + U)(x) = T(x) + U(x)$ for all $x \in V$, and $aT : V \to W$ by $(aT)(x) = aT(x)$ for all $x \in V$.

Theorem 2.7. Let V and W be vector spaces over a field F, and let $T, U: V \to W$ be linear.
   (a) For all $a \in F$, $aT + U$ is linear.
   (b) Using the operations of addition and scalar multiplication in the preceding definition, the collection of all linear transformations from V to W is a vector space over F.

Definitions. Let V and W be vector spaces over F. We denote the vector space of all linear transformations from V into W by $\mathcal{L}(V, W)$. In the case that $V = W$, we write $\mathcal{L}(V)$ instead of $\mathcal{L}(V, W)$.

Theorem 2.8. Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively, and let $T, U: V \to W$ be linear transformations. Then
   (a) $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$ and
   (b) $[aT]_\beta^\gamma = a[T]_\beta^\gamma$ for all scalars a.

2.3   Composition of Linear Transformations and Matrix Multiplication

Definition. Let A, B, and C be sets and $f : A \to B$ and $g : B \to C$ be functions. By following f with g, we obtain a function $g \circ f : A \to C$ called the **composite** of g and f. Thus $(g \circ f)(x) = g(f(x))$ for all $x \in A$.

We use the more convenient notation of UT rather than $U \circ T$ for the composite of linear transformations U and T.

Theorem 2.9. Let V, W, and Z be vector spaces over the same field F, and let $T: V \to W$ and $U: W \to Z$ be linear. Then $UT: V \to Z$ is linear.

Theorem 2.10. Let V be a vector space. Let $T, U_1, U_2 \in L(V)$. Then

   (a) $T(U_1 + U_2) = TU_1 + TU_2$ and $(U_1 + U_2)T = U_1 T + U_2 T$

   (b) $T(U_1 U_2) = (TU_1)U_2$

11

(c) $TI = IT = T$

(d) $a(U_1 U_2) = (aU_1)U_2 = U_1(aU_2)$ *for all scalars a*

Definition. *Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. We define the **product** of A and B, denoted AB, to be the $m \times p$ matrix such that*

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \ for \ 1 \le i \le m, \ 1 \le j \le p$$

If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then $(AB)^t = B^t A^t$.

Theorem 2.11. *Let V, W, and Z be finite-dimensional vector spaces with ordered bases α, β, and γ, respectively. Let T: $V \to W$ and U: $W \to Z$ be linear transformations. Then*

$$[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta$$

Corollary. *Let V be a finite-dimensional vector space with an ordered basis β. Let $T, U \in L(V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.*

Definitions. *We define the **Kronecker delta** δ_{ij} by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \ne j$. The $n \times n$ **identity matrix** I_n is defined by $(I_n)_{ij} = \delta_{ij}$.*

Theorem 2.12. *Let A be an $m \times n$ matrix, B and C be $n \times p$ matrices, and D and E be $q \times m$ matrices. Then*

(a) $A(B + C) = AB + AC$ and $(D + E)A = DA + EA$.

(b) $a(AB)=(aA)B = A(aB)$ for any scalar a.

(c) $I_m A = A = A I_n$.

(d) If V is an n-dimensional vector space with an ordered basis β, then $[I_V]_\beta = I_n$.

Corollary. *Let A be an $m \times n$ matrix, $B_1, B_2, ..., B_k$ be $n \times p$ matrices, $C_1, C_2, ..., C_k$ be $q \times m$ matrices, and $a_1, a_2, ..., a_k$ be scalars. Then*

$$A \left(\sum_{i=1}^{k} a_i B_i \right) = \sum_{i=1}^{k} a_i A B_i$$

and

$$\left(\sum_{i=1}^{k} a_i C_i \right) A = \sum_{i=1}^{n} a_i C_i A$$

Theorem 2.13. *Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. For each $j(1 \le j \le p)$ let u_j and v_j denote the jth columns of AB and B, respectively. Then*

(a) $u_j = A v_j$

(b) $v_j = B e_j$, where e_j is the jth standard vector of F^p.

It follows (see Exercise 14) from Theorem 2.13 that column j of AB is a linear combination of the columns of A with the coefficients in the linear combination being the entries of column j of B. An analogous result holds for rows; that is, row i of AB is a linear combination of the rows of B with the coefficients in the linear combination being the entries of row i of A.

Theorem 2.14. *Let V and W be finite-dimensional vector spaces having ordered bases β and γ, respectively, and let T: $V \to W$ be linear. Then, for each $u \in V$, we have*

$$[T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta$$

Definition. *Let A be an $m \times n$ matrix with entries from a field F. We denote by L_A the mapping $L_A : F_n \to F_m$ defined by $L_A(x) = Ax$ (the matrix product of A and x) for each column vector $x \in F_n$. We call L_A a **left-multiplication transformation**.*

Theorem 2.15. *Let A be an $m \times n$ matrix with entries from F. Then the left-multiplication transformation $L_A : F_n \to F_m$ is linear. Furthermore, if B is any other $m \times n$ matrix (with entries from F) and β and γ are the standard ordered bases for F_n and F_m, respectively, then we have the following properties.*

(a) $[L_A]_\beta^\gamma = A$.

(b) $L_A = L_B$ *if and only if* $A = B$.

(c) $L_{A+B} = L_A + L_B$ *and* $L_{aA} = aL_A$ *for all* $a \in F$.

(d) *If* $T : F^n \to F^m$ *is linear, then there exists a unique $m \times n$ matrix C such that* $T = L_C$. *In fact,* $C = [T]_\beta^\gamma$

(e) *If E is an $n \times p$ matrix, then* $L_{AE} = L_A L_E$.

(f) *If $m = n$, then* $L_{I_n} = I_{F_n}$.

Theorem 2.16. *Let A, B, and C be matrices such that A(BC) is defined. Then (AB)C is also defined and A(BC)=(AB)C; that is, matrix multiplication is associative.*

2.4 Invertibility and Isomorphisms

Definition. *Let V and W be vector spaces, and let T: $V \to W$ be linear. A function U: $W \to V$ is said to be an **inverse** of T if $TU = I_W$ and $UT = I_V$. If T has an inverse, then T is said to be **invertible**. As noted in Appendix B, if T is invertible, then the inverse of T is unique and is denoted by T^{-1}.*

The following facts hold for invertible functions T and U.
1. $(TU)^{-1} = U^{-1}T^{-1}$.
2. $(T^{-1})^{-1} = T$; in particular, T^{-1} is invertible.

We often use the fact that a function is invertible if and only if it is both one-to-one and onto. We can therefore restate Theorem 2.5 as follows.

3. Let T: V → W be a linear transformation, where V and W are finite-dimensional spaces of equal dimension. Then T is invertible if and only if rank(T) = dim(V).

Theorem 2.17. *Let V and W be vector spaces, and let T: V → W be linear and invertible. Then $T^{-1} : W \to V$ is linear.*

Definition. *Let A be an n × n matrix. Then A is invertible if there exists an n × n matrix B such that $AB = BA = I$.*

The matrix B is called the **inverse** of A and is denoted by A^{-1}.

Lemma. *Let T be an invertible linear transformation from V to W. Then V is finite-dimensional if and only if W is finite-dimensional. In this case, dim(V) = dim(W).*

Theorem 2.18. *Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively. Let T: V → W be linear. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible. Furthermore, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.*

Corollary 1. *Let V be a finite-dimensional vector space with an ordered basis β, and let T: V → V be linear. Then T is invertible if and only if $[T]_\beta$ is invertible. Furthermore, $[T^{-1}]_\beta = ([T]_\beta)^{-1}$.*

Corollary 2. *Let A be an n × n matrix. Then A is invertible if and only if L_A is invertible. Furthermore, $(L_A)^{-1} = L_{A^{-1}}$.*

Definitions. *Let V and W be vector spaces. We say that V is **isomorphic** to W if there exists a linear transformation T: V → W that is invertible. Such a linear transformation is called an **isomorphism** from V onto W.*

Theorem 2.19. *Let V and W be finite-dimensional vector spaces (over the same field). Then V is isomorphic to W if and only if dim(V) = dim(W).*

Corollary. *Let V be a vector space over F. Then V is isomorphic to F_n if and only if dim(V) = n.*

Theorem 2.20. *Let V and W be finite-dimensional vector spaces over F of dimensions n and m, respectively, and let β and γ be ordered bases for V and W, respectively. Then the function $\Phi : \mathcal{L}(V, W) \to M_{m \times n}(F)$, defined by $\Phi(T) = [T]_\beta^\gamma$ for $T \in L(V, W)$, is an isomorphism.*

Corollary. *Let V and W be finite-dimensional vector spaces of dimensions n and m, respectively. Then $\mathcal{L}(V, W)$ is finite-dimensional of dimension mn.*

Definition. *Let β be an ordered basis for an n-dimensional vector space V over the field F. The **standard representation of** V **with respect to** β is the function $\phi_\beta : V \to F^n$ defined by $\phi_\beta(x) = [x]_\beta$ for each $x \in V$.*

Theorem 2.21. *For any finite-dimensional vector space V with ordered basis β, ϕ_β is an isomorphism.*

2.5 The Change of Coordinate Matrix

Theorem 2.22. *Let β and β' be two ordered bases for a finite-dimensional vector space V, and let $Q = [I_V]_{\beta'}^{\beta}$. Then*
    (a) Q is invertible.
    (b) For any $v \in V$, $[v]_\beta = Q[v]_{\beta'}$.

The matrix $Q = [I_V]_{\beta'}^{\beta}$ defined in Theorem 2.22 is called a **change of coordinate matrix**. Because of part (b) of the theorem, we say that Q **changes β'-coordinates into β-coordinates**. Observe that if $\beta = \{x_1, x_2, \ldots, x_n\}$ and $\beta' = \{x_1', x_2', \ldots, x_n'\}$ then

$$x_j' = \sum_{i=1}^{n} Q_{ij} x_i$$

for j = 1, 2,... ,n; that is, the jth column of Q is $[x_j']_\beta$.

Notice that if Q changes β'-coordinates into β-coordinates, then Q^{-1} changes β-coordinates into β'-coordinates.

For the remainder of this section, we consider only linear transformations that map a vector space V into itself. Such a linear transformation is called a **linear operator** on V.

Theorem 2.23. *Let T be a linear operator on a finite-dimensional vector space V, and let β and β' be ordered bases for V. Suppose that Q is the change of coordinate matrix that changes β'-coordinates into β-coordinates. Then*

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q$$

Corollary. *Let $A \in M_{n \times n}(F)$, and let γ be an ordered basis for F^n. Then $[L_A]\gamma = Q^{-1}AQ$, where Q is the $n \times n$ matrix whose jth column is the jth vector of γ.*

Definition. *Let A and B be matrices in $M_{n \times n}(F)$. We say that B is **similar** to A if there exists an invertible matrix Q such that $B = Q^{-1}AQ$.*

2.6 Dual Spaces*

In this section, we are concerned exclusively with linear transformations from a vector space V into its field of scalars F, which is itself a vector space of dimension 1 over F. Such a linear transformation is called a **linear functional** on V. We generally use the letters f, g, h,... to denote linear functionals.

Let V be a finite-dimensional vector space, and let $\beta = \{x_1, x_2, \ldots, x_n\}$ be an ordered basis for V. For each i = 1, 2,...,n, define $f_i(x) = a_i$, where

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

is the coordinate vector of x relative to β. Then f_i is a linear functional on V called the **ith coordinate function with respect to the basis** β. Note that $f_i(x_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta.

Definition. *For a vector space V over F, we define the **dual space** of V to be the vector space $\mathcal{L}(V, F)$, denoted by V^*.*

Hence by Theorem 2.19, V and V^* are isomorphic. We also define the **double dual** V^{**} of V to be the dual of V^*.

Theorem 2.24. *Suppose that V is a finite-dimensional vector space with the ordered basis $\beta = \{x_1, x_2, \ldots, x_n\}$. Let $f_i(1 \leq i \leq n)$ be the ith coordinate function with respect to β as just defined, and let $\beta^* = \{f_1, f_2, \ldots, f_n\}$. Then β^* is an ordered basis for V^*, and, for any $f \in V^*$, we have*

$$f = \sum_{i=1}^{n} f(x_i) f_i$$

Definition. *Using the notation of Theorem 2.24, we call the ordered basis $\beta^* = \{f_1, f_2, \ldots, f_n\}$ of V^* that satisfies $f_i(x_j) = \delta_{ij}(1 \leq i, j \leq n)$ the **dual basis** of β.*

Theorem 2.25. *Let V and W be finite-dimensional vector spaces over F with ordered bases β and γ, respectively. For any linear transformation T: $V \rightarrow W$, the mapping $T^t : W^* \rightarrow V^*$ defined by $T^t(g) = gT$ for all $g \in W^*$ is a linear transformation with the property that $[T^t]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^t$.*

The linear transformation T^t defined in Theorem 2.25 is called the **transpose** of T. It is clear that T^t is the unique linear transformation U such that $[U]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^t$.

For a vector $x \in V$, we define $\hat{x} : V^* \rightarrow F$ by $\hat{x}(f) = f(x)$ for every $f \in V^*$. It is easy to verify that $\hat{x}$ is a linear functional on V^*, so $\hat{x} \in V^{**}$.

Lemma. *Let V be a finite-dimensional vector space, and let $x \in V$. If $\hat{x}(f) = 0$ for all $f \in V^*$, then $x = 0$.*

Theorem 2.26. *Let V be a finite-dimensional vector space, and define $\Psi : V \rightarrow V^{**}$ by $\Psi(x) = \hat{x}$. Then Ψ is an isomorphism.*

Corollary. *Let V be a finite-dimensional vector space with dual space V^*. Then every ordered basis for V^* is the dual basis for some basis for V.*

Definition. *V denotes a finite-dimensional vector space over F. For every subset S of V, define the **annihilator** S^0 of S as*

$$S^0 = \{f \in V^* : f(x) = 0 \ for \ all \ x \ \in S\}$$

2.7 Homogeneous Linear Differential Equations with Constant Coefficients*

Gugugu.

3 Elementary Matrix Operations and Systems of Linear Equations

3.1 Elementary Matrix Operations and Elementary Matrices

Definitions. *Let A be an $m \times n$ matrix. Any one of the following three operations on the rows [columns] of A is called an **elementary row [column] operation**:*

(1) interchanging any two rows [columns] of A;

(2) multiplying any row [column] of A by a nonzero scalar;

(3) adding any scalar multiple of a row [column] of A to another row [column].

Any of these three operations is called an **elementary operation**. Elementary operations are of **type 1**, **type 2**, or **type 3** depending on whether they are obtained by (1), (2), or (3).

Definition. *An $n \times n$ **elementary matrix** is a matrix obtained by performing an elementary operation on I_n. The elementary matrix is said to be of **type 1**, **2**, or **3** according to whether the elementary operation performed on In is a type 1, 2, or 3 operation, respectively.*

Theorem 3.1. *Let $A \in M_{m \times n}(F)$, and suppose that B is obtained from A by performing an elementary row [column] operation. Then there exists an $m \times m$ [$n \times n$] elementary matrix E such that $B = EA$ [$B = AE$]. In fact, E is obtained from $I_m[I_n]$ by performing the same elementary row [column] operation as that which was performed on A to obtain B. Conversely, if E is an elementary $m \times m$ [$n \times n$] matrix, then EA [AE] is the matrix obtained from A by performing the same elementary row [column] operation as that which produces E from Im [In].*

Theorem 3.2. *Elementary matrices are invertible, and the inverse of an elementary matrix is an elementary matrix of the same type.*

3.2 The Rank of a Matrix and Matrix Inverses

Definition. *If $A \in M_{m \times n}(F)$, we define the **rank** of A, denoted rank(A), to be the rank of the linear transformation $L_A : F_n \to F_m$.*

Many results about the rank of a matrix follow immediately from the corresponding facts about a linear transformation. An important result of this type, which follows from Fact 3 (p. 100) and Corollary 2 to Theorem 2.18 (p. 102), is that an $n \times n$ matrix is invertible if and only if its rank is n.

Theorem 3.3. *Let $T : V \to W$ be a linear transformation between finitedimensional vector spaces, and let β and γ be ordered bases for V and W, respectively. Then $rank(T) = rank([T]_\beta^\gamma)$.*

Theorem 3.4. *Let A be an m × n matrix. If P and Q are invertible m × m and n × n matrices, respectively, then*

(a) *rank(AQ) = rank(A),*

(b) *rank(PA) = rank(A), and therefore,*

(c) *rank(PAQ) = rank(A).*

Corollary. *Elementary row and column operations on a matrix are rankpreserving.*

Theorem 3.5. *The rank of any matrix equals the maximum number of its linearly independent columns; that is, the rank of a matrix is the dimension of the subspace generated by its columns.*

Theorem 3.6. *Let A be an m × n matrix of rank r. Then $r \leq m, r \leq n$, and, by means of a finite number of elementary row and column operations, A can be transformed into the matrix*

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

, where O_1, O_2, and O_3 are zero matrices. Thus $D_{ii} = 1$ for $i \leq r$ and $D_{ij} = 0$ otherwise.

Corollary 1. *Let A be an m × n matrix of rank r. Then there exist invertible matrices B and C of sizes m × m and n×n, espectively, such that $D = BAC$, where*

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

is the m × n matrix in which O_1, O_2, and O_3 are zero matrices.

Corollary 2. *Let A be an m × n matrix. Then*

(a) *$rank(A^t) = rank(A)$.*

(b) *The rank of any matrix equals the maximum number of its linearly independent rows; that is, the rank of a matrix is the dimension of the subspace generated by its rows.*

(c) *The rows and columns of any matrix generate subspaces of the same dimension, numerically equal to the rank of the matrix.*

Corollary 3. *Every invertible matrix is a product of elementary matrices.*

Theorem 3.7. *Let T: V → W and U: W → Z be linear transformations on finite-dimensional vector spaces V, W, and Z, and let A and B be matrices such that the product AB is defined. Then*

(a) $rank(UT) \leq rank(U)$.

(b) $rank(UT) \leq rank(T)$.

(c) $rank(AB) \leq rank(A)$.

(d) $rank(AB) \leq rank(B)$.

The Inverse of a Matrix

Definition. *Let A and B be m × n and m × p matrices, respectively. By the* ***augmented matrix*** *(A|B), we mean the m × (n + p) matrix (A B), that is, the matrix whose first n columns are the columns of A, and whose last p columns are the columns of B.*

Being able to test for invertibility and compute the inverse of a matrix allows us, with the help of Theorem 2.18 (p. 101) and its corollaries, to test for invertibility and compute the inverse of a linear transformation. The next example demonstrates this technique.

3.3   Systems of Linear Equations - Theoretical Aspects

The system of equations

$$(S) \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

where a_{ij} and b_i ($1 \leq i \leq m$ and $1 \leq j \leq n$) are scalars in a field F and $x_1, x_2, \ldots, x_n$ are n variables taking values in F, is called a **system of** m **linear equations in** n **unknowns over the field** F.

The $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is called the **coefficient matrix** of the system (S).

If we let

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad and \ b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

then the system (S) may be rewritten as a single matrix equation

$$Ax = b$$

A **solution** to the system (S) is an n-tuple

$$s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \in F^n$$

such that $As = b$. The set of all solutions to the system (S) is called the **solution set** of the system. System (S) is called **consistent** if its solution set is nonempty; otherwise it is called **inconsistent**.

We see that a system of linear equations can have one, many, or no solutions.

Definitions. *A system $Ax = b$ of m linear equations in n unknowns is said to be* ***homogeneous*** *if $b = 0$. Otherwise the system is said to be* ***nonhomogeneous***.

Theorem 3.8. *Let $Ax = 0$ be a homogeneous system of m linear equations in n unknowns over a field F. Let K denote the set of all solutions to $Ax = 0$. Then $K = N(L_A)$; hence K is a subspace of F^n of dimension $n - rank(L_A) = n - rank(A)$.*

Corollary. *If m<n, the system $Ax = 0$ has a nonzero solution.*

We refer to the equation Ax = 0 as the **homogeneous system corresponding to** Ax = b.

Theorem 3.9. *Let K be the solution set of a system of linear equations $Ax = b$, and let K_H be the solution set of the corresponding homogeneous system Ax = 0 . Then for any solution s to $Ax = b$*

$$K = \{s\} + K_H = \{s + k : k \in K_H\}$$

Theorem 3.10. *Let $Ax = b$ be a system of n linear equations in n unknowns. If A is invertible, then the system has exactly one solution, namely, $A^{-1}b$. Conversely, if the system has exactly one solution, then A is invertible.*

The matrix (A|b) is called the **augmented matrix** of the system Ax = b.

Theorem 3.11. *Let $Ax = b$ be a system of linear equations. Then the system is consistent if and only if rank(A) = rank(A|b).*

Theorem 3.12.

3.4 Systems of Linear Equations - Computational Aspects

Definition. *Two systems of linear equations are called equivalent if they have the same solution set.*

Theorem 3.13. *Let $Ax = b$ be a system of m linear equations in n unknowns, and let C be an invertible $m \times m$ matrix. Then the system $(CA)x = Cb$ is equivalent to $Ax = b$.*

Corollary. *Let $Ax = b$ be a system of m linear equations in n unknowns. If $(A'|b')$ is obtained from $(A|b)$ by a finite number of elementary row operations, then the system $A'x = b'$ is equivalent to the original system.*

Definition. *A matrix is said to be in **reduced row echelon form** if the following three conditions are satisfied.*
(a) Any row containing a nonzero entry precedes any row in which all the entries are zero (if any).
(b) The first nonzero entry in each row is the only nonzero entry in its column.
(c) The first nonzero entry in each row is 1 and it occurs in a column to the right of the first nonzero entry in the preceding row.

It can be shown (see the corollary to Theorem 3.16) that the reduced row echelon form of a matrix is unique.

The procedure described on pages 183–185 for reducing an augmented matrix to reduced row echelon form is called **Gaussian elimination**. It consists of two separate parts.

1. In the forward pass (steps 1-5), the augmented matrix is transformed into an upper triangular matrix in which the first nonzero entry of each row is 1, and it occurs in a column to the right of the first nonzero entry of each preceding row.

2. In the backward pass or back-substitution (steps 6-7), the upper triangular matrix is transformed into reduced row echelon form by making the first nonzero entry of each row the only nonzero entry of its column.

Theorem 3.14. *Gaussian elimination transforms any matrix into its reduced row echelon form.*

To solve a system for which the augmented matrix is in reduced row echelon form, divide the variables into two sets. The first set consists of those variables that appear as leftmost variables in one of the equations of the system (in this case the set is $\{x_1, x_2, x_4\}$). The second set consists of all the remaining variables (in this case, $\{x_3, x_5\}$). To each variable in the second set, assign a parametric value $t_1, t_2, \ldots (x_3 = t_1, x_5 = t_2)$, and then solve for the variables of the first set in terms of those in the second set. See page 188 of the textbook.

Theorem 3.15. *Let $Ax = b$ be a system of r nonzero equations in n unknowns. Suppose that rank(A) = rank(A|b) and that (A|b) is in reduced row echelon form. Then*

(a) rank(A) = r.

(b) If the general solution obtained by the procedure above is of the form

$$s = s_0 + t_1 u_1 + t_2 u_2 + \cdots + t_{n-r} u_{n-r},$$

then $\{u_1, u_2, \ldots, u_{n-r}\}$ is a basis for the solution set of the corresponding homogeneous system, and s0 is a solution to the original system.

Theorem 3.16. *Let A be an $m \times n$ matrix of rank r, where $r > 0$, and let B be the reduced row echelon form of A. Then*

(a) The number of nonzero rows in B is r.

(b) For each i = 1, 2, . . . , r, there is a column b_{j_i} of B such that $b_{j_i} = e_i$.

(c) The columns of A numbered $j_1, j_2, \ldots, j_r$ are linearly independent.

(d) For each k = 1, 2,...n, if column k of B is $d_1 e_1 + d_2 e_2 + \cdots + d_r e_r$, then column k of A is $d_1 a_{j_1} + d_2 a_{j_2} + \cdots + d_r a_{j_r}$.

Corollary. *The reduced row echelon form of a matrix is unique.*

4 Determinants

4.1 Determinants of Order 2

Definition. *If*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

*is a 2×2 matrix with entries from a field F, then we define the **determinant** of A, denoted $\det(A)$ or $|A|$, to be the scalar $ad - bc$.*

Since $\det(A + B) \neq \det(A) + \det(B)$, the function $\det : M_{2\times 2}(R) \to R$ is not a linear transformation.

Theorem 4.1. *The function $\det : M_{2\times 2}(F) \to F$ is a linear function of each row of a 2×2 matrix when the other row is held fixed. That is, if u, v, and w are in F^2 and k is a scalar, then*

$$\det \begin{pmatrix} u + kv \\ w \end{pmatrix} = \det \begin{pmatrix} u \\ w \end{pmatrix} + k \det \begin{pmatrix} v \\ w \end{pmatrix}$$

and

$$\det \begin{pmatrix} w \\ u + kv \end{pmatrix} = \det \begin{pmatrix} w \\ u \end{pmatrix} + k \det \begin{pmatrix} w \\ v \end{pmatrix}$$

Theorem 4.2. *Let $A \in M_{2\times 2}(F)$. Then the determinant of A is nonzero if and only if A is invertible. Moreover, if A is invertible, then*

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

The Area of a Parallelogram By the **angle** between two vectors in R^2, we mean the angle with measure $\theta(0 \leq \theta < \pi)$ that is formed by the vectors having the same magnitude and direction as the given vectors but emanating from the origin.

If $\beta = \{u, v\}$ is an ordered basis for R^2, we define the orientation of β to be the real number

$$O\begin{pmatrix} u \\ v \end{pmatrix} = \frac{\det \begin{pmatrix} u \\ v \end{pmatrix}}{\left| \det \begin{pmatrix} u \\ v \end{pmatrix} \right|}$$

$$O\begin{pmatrix} u \\ v \end{pmatrix} = 1$$

if and only if the ordered basis $\{u, v\}$ forms a **right-handed** coordinate system. Otherwise, $\{u, v\}$ is called a **left-handed** system and

$$O\begin{pmatrix} u \\ v \end{pmatrix} = -1$$

For convenience, we also define

$$O\begin{pmatrix} u \\ v \end{pmatrix} = 1$$

if $\{u, v\}$ is linearly dependent.

4.2 Determinants of Order n

Given $A \in M_{n \times n}(F)$, for $n \geq 2$, denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j by $\tilde{A}_{ij}$.

Definitions. *Let $A \in M_{n \times n}(F)$. If $n = 1$, so that $A = (A_{11})$, we define $\det(A) = A_{11}$. For $n \geq 2$, we define $\det(A)$ recursively as*

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{A}_{1j})$$

*The scalar $\det(A)$ is called the **determinant** of A and is also denoted by $|A|$.*
The scalar

$$(-1)^{i+j} \det(\tilde{A}_{ij})$$

*is called the **cofactor** of the entry of A in row i, column j.*

Thus the determinant of A equals the sum of the products of each entry in row 1 of A multiplied by its cofactor. This formula is called **cofactor expansion along the first row** of A.

Theorem 4.3. *The determinant of an $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed. That is, for $1 \leq r \leq n$, we have*

$$\det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u + kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ ar + 1 \\ \vdots \\ a_n \end{pmatrix} + k \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

whenever k is a scalar and u, v, and each a_i are row vectors in F^n.

Corollary. *If $A \in M_{n \times n}(F)$ has a row consisting entirely of zeros, then $\det(A) = 0$.*

Lemma. *Let $B \in M_{n \times n}(F)$, where $n \geq 2$. If row i of B equals e_k for some $k(1 \leq k \leq n)$, then $\det(B) = (-1)^{i+k} \det(\tilde{B}_{ik})$.*

Theorem 4.4. *The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is, if $A \in M_{n \times n}(F)$, then for any integer $i(1\ i\ n)$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij})$$

Corollary. *If $A \in M_{n \times n}(F)$ has two identical rows, then $\det(A) = 0$.*

Theorem 4.5. *If $A \in M_{n \times n}(F)$ and B is a matrix obtained from A by interchanging any two rows of A, then $\det(B) = -\det(A)$.*

Theorem 4.6. *Let $A \in M_{n \times n}(F)$, and let B be a matrix obtained by adding a multiple of one row of A to another row of A. Then $\det(B) = \det(A)$.*

Corollary. *If $A \in M_{n \times n}(F)$ has rank less than n, then $\det(A) = 0$.*

The following rules summarize the effect of an elementary row operation on the determinant of a matrix $A \in M_{n \times n}(F)$.

(a) If B is a matrix obtained by interchanging any two rows of A, then $\det(B) = -\det(A)$.

(b) If B is a matrix obtained by multiplying a row of A by a nonzero scalar k, then $\det(B) = k \det(A)$.

(c) If B is a matrix obtained by adding a multiple of one row of A to another row of A, then $\det(B) = \det(A)$.

The determinant of an upper triangular matrix is the product of its diagonal entries.

4.3 Properties of Determinants

Because the determinant of the n×n identity matrix is 1, we can interpret the statements on page 217 as the following facts about the determinants of elementary matrices.

(a) If E is an elementary matrix obtained by interchanging any two rows of I, then $\det(E) = -1$.

(b) If E is an elementary matrix obtained by multiplying some row of I by the nonzero scalar k, then $\det(E) = k$.

(c) If E is an elementary matrix obtained by adding a multiple of some row of I to another row, then $\det(E) = 1$.

Theorem 4.7. *For any $A, B \in M_{n \times n}(F)$, $\det(AB) = \det(A) \cdot \det(B)$.*

Corollary. *A matrix $A \in M_{n \times n}(F)$ is invertible if and only if $\det(A) \neq 0$. Furthermore, if A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.*

Theorem 4.8. *For any $A \in M_{n \times n}(F)$, $\det(A^t) = \det(A)$.*

Theorem 4.9 (Cramer's Rule)**.** *Let $Ax = b$ be the matrix form of a system of n linear equations in n unknowns, where $x = (x_1, x_2, \ldots, x_n)^t$. If $\det(A) \neq 0$, then this system has a unique solution, and for each $k(k = 1, 2, \ldots, n)$,*

$$x_k = \frac{\det(M_k)}{\det(A)}$$

where M_k is the $n \times n$ matrix obtained from A by replacing column k of A by b.

If the rows of A are $a_1, a_2, \ldots, a_n$, respectively, then $|\det(A)|$ is the **n-dimensional volume** (the generalization of area in R^2 and volume in R^3) of the parallelepiped having the vectors $a_1, a_2, \ldots, a_n$ as adjacent sides.

Specifically, if γ is any ordered basis for R^n and β is the standard ordered basis for R^n, then γ induces a right-handed coordinate system if and only if $\det(Q) > 0$, where Q is the change of coordinate matrix changing γ-coordinates into β-coordinates.

More generally, if β and γ are two ordered bases for R^n, then the coordinate systems induced by β and γ have the same **orientation** (either both are right-handed or both are left-handed) if and only if $\det(Q) > 0$, where Q is the change of coordinate matrix changing γ-coordinates into β-coordinates.

4.4 Summary - Important Facts about Determinants

If A and B are similar matrices, then $\det(A) = \det(B)$.

4.5 A Characterization of the Determinant*

5 Diagonalization

5.1 Eigenvalues and Eigenvectors

Definitions. *A linear operator T on a finite-dimensional vector space V is called* ***diagonalizable*** *if there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix. A square matrix A is called diagonalizable if L_A is* ***diagonalizable***.

Definitions. *Let T be a linear operator on a vector space V. A nonzero vector $v \in V$ is called an* ***eigenvector*** *of T if there exists a scalar λ such that $T(v) = \lambda v$. The scalar λ is called the* ***eigenvalue*** *corresponding to the eigenvector v.*

Let A be in $M_{n \times n}(F)$. A nonzero vector $v \in F^n$ is called an ***eigenvector*** *of A if v is an eigenvector of L_A; that is, if $Av = \lambda v$ for some scalar λ. The scalar λ is called the* ***eigenvalue*** *of A corresponding to the eigenvector v.*

Theorem 5.1. *A linear operator T on a finite-dimensional vector space V is diagonalizable if and only if there exists an ordered basis β for V consisting of eigenvectors of T. Furthermore, if T is diagonalizable, $\beta = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of eigenvectors of T, and $D = [T]_\beta$, then D is a diagonal matrix and D_{jj} is the eigenvalue corresponding to v_j for $1 \le j \le n$.*

Theorem 5.2. *Let $A \in M_{n \times n}(F)$. Then a scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.*

Definition. *Let $A \in M_{n \times n}(F)$. The polynomial $f(t) = \det(A - tI_n)$ is called the* ***characteristic polynomial*** *of A.*

Theorem 5.2 states that the eigenvalues of a matrix are the zeros of its characteristic polynomial. When determining the eigenvalues of a matrix or a linear operator, we normally compute its characteristic polynomial, as in the next example.

Definition. *Let T be a linear operator on an n-dimensional vector space V with ordered basis β. We define the characteristic polynomial $f(t)$ of T to be the* ***characteristic polynomial*** *of $A = [T]_\beta$. That is,*

$$f(t) = \det(A - tI_n)$$

The remark preceding this definition shows that the definition is independent of the choice of ordered basis . We often denote the characteristic polynomial of an operator T by $\det(T - tI)$.

Theorem 5.3. *Let $A \in M_{n \times n}(F)$.*

(a) The characteristic polynomial of A is a polynomial of degree n with leading coefficient $(-1)^n$.

(b) A has at most n distinct eigenvalues.

Theorem 5.4. *Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. A vector $v \in V$ is an eigenvector of T corresponding to λ if and only if $v \neq 0$ and $v \in N(T - \lambda I)$.*

To find the eigenvectors of a linear operator T on an n-dimensional vector space, select an ordered basis β for V and let $A = [T]_\beta$. $v \in V$ is an eigenvector of T corresponding to λ if and only if $\phi_\beta(v)$ is an eigenvector of A corresponding to λ.

5.2 Diagonalizability

Theorem 5.5. *Let T be a linear operator on a vector space V, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of T. If $v_1, v_2, \ldots, v_k$ are eigenvectors of T such that λ_i corresponds to $v_i (1 \leq i \leq k)$, then $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.*

Corollary. *Let T be a linear operator on an n-dimensional vector space V. If T has n distinct eigenvalues, then T is diagonalizable.*

Definition. *A polynomial f(t) in P(F) splits over F if there are scalars $c, a_1, \ldots, a_n$ (not necessarily distinct) in F such that*

$$f(t) = c(t - a_1)(t - a_2) \cdots (t - a_n)$$

Theorem 5.6. *The characteristic polynomial of any diagonalizable linear operator splits.*

Definition. *Let λ be an eigenvalue of a linear operator or matrix with characteristic polynomial f(t). The **(algebraic) multiplicity** of λ is the largest positive integer k for which $(t - \lambda)^k$ is a factor of f(t).*

Definition. *Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. Define $E_\lambda = \{x \in V : T(x) = \lambda x\} = N(T - \lambda I_V)$. The set E_λ is called the **eigenspace** of T corresponding to the eigenvalue λ. Analogously, we define the eigenspace of a square matrix A to be the **eigenspace** of L_A.*

Theorem 5.7. *Let T be a linear operator on a finite-dimensional vector space V, and let λ be an eigenvalue of T having multiplicity m. Then $1 \leq \dim(E_\lambda) \leq m$.*

Lemma. *Let T be a linear operator, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of T. For each $i = 1, 2, \ldots, k$, let $v_i \in E_{\lambda_i}$, the eigenspace corresponding to λ_i. If*

$$v_1 + v_2 + \cdots + v_k = 0$$

then $v_i = 0$ for all i.

Theorem 5.8. *Let T be a linear operator on a vector space V, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of T. For each $i = 1, 2, \ldots, k$, let S_i be a finite linearly independent subset of the eigenspace E_{λ_i}. Then $S = S_1 \cup S_2 \cup \cdots \cup S_k$ is a linearly independent subset of V.*

Theorem 5.9. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then*

(a) T is diagonalizable if and only if the multiplicity of λ_i is equal to $\dim(E_{\lambda_i})$ for all i.

(b) If T is diagonalizable and β_i is an ordered basis for E_{λ_i} for each i, then $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is an ordered basis for V consisting of eigenvectors of T.

Test for Diagonalization Let T be a linear operator on an n-dimensional vector space V. Then T is diagonalizable if and only if both of the following conditions hold.

1. The characteristic polynomial of T splits.

2. For each eigenvalue λ of T, the multiplicity of λ equals $n - rank(T - \lambda I)$.

These same conditions can be used to test if a square matrix A is diagonalizable because diagonalizability of A is equivalent to diagonalizability of the operator L_A.

Systems of Differential Equations Use diagonalization to reformulate a differential equation system to many differential equations.

Direct Sums*

Definition. *Let $W_1, W_2, \ldots, W_k$ be subspaces of a vector space V. We define the **sum** of these subspaces to be the set*

$$\{v_1 + v_2 + \cdots + v_k : v_i \in W_i \text{ for } 1 \le i \le k\}$$

which we denote by $W_1 + W_2 + \cdots + W_k$ or $\sum_{i=1}^{k} W_i$.

It is a simple exercise to show that the sum of subspaces of a vector space is also a subspace.

Definition. *Let $W_1, W_2, \ldots, W_k$ be subspaces of a vector space V. We call V the **direct sum** of the subspaces $W_1, W_2, \ldots, W_k$ and write $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$, if*

$$V = \sum_{i=1}^{k} W_i$$

and

$$W_j \cap \sum_{i \ne j} W_i = \{0\} \text{ for each } j (1 \le j \le k)$$

Theorem 5.10. *Let $W_1, W_2, \ldots, W_k$ be subspaces of a finite-dimensional vector space V. The following conditions are equivalent.*

(a) $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

28

(b) $V = \sum_{i=1}^{k} W_i$ and, for any vectors $v_1, v_2, \ldots, v_k$ such that $v_i \in W_i (1 \leq i \leq k)$, if $v_1 + v_2 + \cdots + v_k = 0$, then $v_i = 0$ for all i.

(c) Each vector $v \in V$ can be uniquely written as $v = v_1 + v_2 + \cdots + v_k$, where $v_i \in W_i$.

(d) If γ_i is an ordered basis for $W_i (1 \leq i \leq k)$, then $\gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_k$ is an ordered basis for V.

(e) For each $i = 1, 2, ..., k$, there exists an ordered basis γ_i for W_i such that $\gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_k$ is an ordered basis for V.

Theorem 5.11. *A linear operator T on a finite-dimensional vector space V is diagonalizable if and only if V is the direct sum of the eigenspaces of T.*

5.3   Matrix Limits and Markov Chains*

Gugugu

Theorem 5.12.

Theorem 5.13.

Theorem 5.14.

Theorem 5.15.

Theorem 5.16.

Theorem 5.17.

Theorem 5.18.

Theorem 5.19.

Theorem 5.20.

5.4   Invariant Subspaces and the Cayley-Hamilton Theorem

Definition. *Let T be a linear operator on a vector space V. A subspace W of V is called a T-**invariant subspace** of V if $T(W) \subseteq W$, that is, if $T(v) \in W$ for all $v \in W$.*

Let T be a linear operator on a vector space V, and let x be a nonzero vector in V. The subspace

$$W = span(\{x, T(x), T^2(x), \ldots\})$$

is called the T-**cyclic subspace** of V **generated by** x. It is a simple matter to show that W is T-invariant. In fact, W is the "smallest" T-invariant subspace of V containing x. That is, any T-invariant subspace of V containing x must also contain W.

Theorem 5.21. *Let T be a linear operator on a finite-dimensional vector space V, and let W be a T-invariant subspace of V. Then the characteristic polynomial of TW divides the characteristic polynomial of T.*

Theorem 5.22. *Let T be a linear operator on a finite-dimensional vector space V, and let W denote the T-cyclic subspace of V generated by a nonzero vector $v \in V$. Let $k = dim(W)$. Then*

(a) $\{v, T(v), T^2(v), \ldots, T^{k-1}(v)\}$ *is a basis for W.*

(b) *If $a_0 v + a_1 T(v) + \cdots + a_{k-1} T^{k-1}(v) + T^k(v) = 0$, then the characteristic polynomial of T_W is $f(t) = (-1)^k (a_0 + a_1 t + \cdots + a_{k-1} t^{k-1} + t^k)$.*

The Cayley–Hamilton Theorem

Theorem 5.23 (Cayley–Hamilton)**.** *Let T be a linear operator on a finite-dimensional vector space V, and let f(t) be the characteristic polynomial of T. Then $f(T) = T_0$, the zero transformation. That is, T "satisfies" its characteristic equation.*

Corollary (Cayley–Hamilton Theorem for Matrices)**.** *Let A be an $n \times n$ matrix, and let f(t) be the characteristic polynomial of A. Then f(A) = O, the $n \times n$ zero matrix.*

Invariant Subspaces and Direct Sums

Theorem 5.24. *Let T be a linear operator on a finite-dimensional vector space V, and suppose that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$, where W_i is a T-invariant subspace of V for each i $(1 \leq i \leq k)$. Suppose that $f_i(t)$ is the characteristic polynomial of T_{W_i} $(1 \leq i \leq k)$. Then $f_1(t) \cdot f_2(t) \cdot \cdots \cdot f_k(t)$ is the characteristic polynomial of T.*

Definition. *Let $B_1 \in M_{m \times m}(F)$, and let $B_2 \in M_{n \times n}(F)$. We define the **direct sum** of B_1 and B_2, denoted $B_1 \oplus B_2$, as the $(m+n) \times (m+n)$ matrix A such that*

$$
A_{ij} = \begin{cases} (B_1)_{ij} & for \ 1 \leq i,j \leq m \\ (B_2)_{(i-m),(j-m)} & for \ m+1 \leq i,j \leq n+m \\ 0 & otherwise. \end{cases}
$$

*If $B_1, B_2, \ldots, B_k$ are square matrices with entries from F, then we define the **direct sum** of $B_1, B_2, \ldots, B_k$ recursively by*

$$
B_1 \oplus B_2 \oplus \cdots \oplus B_k = (B_1 \oplus B_2 \oplus \cdots \oplus B_{k-1}) \oplus B_k
$$

If $A = B_1 \oplus B_2 \oplus \cdots \oplus B_k$, then we often write

$$
A = \begin{pmatrix} B_1 & O & \cdots & O \\ O & B_2 & \cdots & O \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & B_k \end{pmatrix}
$$

Theorem 5.25. *Let T be a linear operator on a finite-dimensional vector space V, and let $W_1, W_2, ..., W_k$ be T-invariant subspaces of V such that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. For each i, let β_i be an ordered basis for W_i, and let $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$. Let $A = [T]_\beta$ and $B_i = [T_{W_i}]_{\beta_i}$ for $i = 1, 2, ..., k$. Then $A = B_1 \oplus B_2 \oplus \cdots \oplus B_k$.*

Definition. *Let T be a linear operator on a vector space V, and let W be a T-invariant subspace of V. Define $\overline{T} : V/W \to V/W$ by*

$$\overline{T}(v + W) = T(v) + W \ \text{for any } v + W \in V/W$$

6 Inner Product Spaces

6.1 Inner Products and Norms

Definition. *Let V be a vector space over F. An **inner product** on V is a function that assigns, to every ordered pair of vectors x and y in V, a scalar in F, denoted $\langle x, y \rangle$, such that for all x, y, and z in V and all c in F, the following hold:*

(a) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$.

(b) $\langle cx, y \rangle = c \langle x, y \rangle$.

(c) $\overline{\langle x, y \rangle} = \langle y, x \rangle$, *where the bar denotes complex conjugation.*

(d) $\langle x, x \rangle > 0$ *if $x \neq 0$.*

For $x = (a_1, a_2, ..., a_n)$ and $y = (b_1, b_2, ..., b_n)$ in F_n, define

$$\langle x, y \rangle = \sum_{i=1}^{n} a_i \overline{b_i}$$

The inner product above is called the **standard inner product** on F^n. When F $=$ R the conjugations are not needed, and in early courses this standard inner product is usually called the *dot product* and is denoted by $x \cdot y$ instead of $\langle x, y \rangle$.

Definition. *Let $A \in M_{m \times n}(F)$. We define the **conjugate transpose** or **adjoint** of A to be the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{A_{ji}}$ for all i, j.*

Notice that if x and y are viewed as column vectors in F^n, then $\langle x, y \rangle = y^* x$.

The inner product on $M_{n \times n}(F)$ defined by $\langle A, B \rangle = tr(B^* A)$ is called the **Frobenius inner product**.

A vector space V over F endowed with a specific inner product is called an **inner product space**. If F $=$ C, we call V a **complex inner product space**, whereas if F $=$ R, we call V a **real inner product space**.

For the remainder of this chapter, F^n denotes the inner product space with the standard inner product. Likewise, $M_{n \times n}(F)$ denotes the inner product space with the Frobenius inner product.

The reader is cautioned that two distinct inner products on a given vector space yield two distinct inner product spaces.

A very important inner product space that resembles $C([0,1])$ is the space H of continuous complex-valued functions defined on the interval $[0, 2\pi]$ with the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) \overline{g(t)} \mathrm{d}t$$

The reason for the constant $1/2\pi$ will become evident later. This inner product space, which arises often in the context of physical situations, is examined more closely in later sections.

Theorem 6.1. *Let V be an inner product space. Then for $x, y, z \in V$ and $c \in F$, the following statements are true.*

(a) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.

(b) $\langle x, cy \rangle = \overline{c} \langle x, y \rangle$.

(c) $\langle x, 0 \rangle = \langle 0, x \rangle = 0$.

(d) $\langle x, x \rangle = 0$ if and only if $x = 0$.

(e) If $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in V$, then $y = z$.

The reader should observe that (a) and (b) of Theorem 6.1 show that the inner product is **conjugate linear** in the second component.

Definition. *Let V be an inner product space. For $x \in V$, we define the **norm** or **length** of x by $\|x\| = \sqrt{\langle x, x \rangle}$.*

Theorem 6.2. *Let V be an inner product space over F. Then for all $x, y \in V$ and $c \in F$, the following statements are true.*

(a) $\|cx\| = \|c\| \cdot \|x\|$.

(b) $\|x\| = 0$ if and only if $x = 0$. In any case, $\|x\| \geq 0$.

(c) (Cauchy-Schwarz Inequality) $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$.

(d) (Triangle Inequality) $\|x + y\| \leq \|x\| + \|y\|$.

Definition. *Let V be an inner product space. Vectors x and y in V are **orthogonal (perpendicular)** if $\langle x, y \rangle = 0$. A subset S of V is **orthogonal** if any two distinct vectors in S are orthogonal. A vector x in V is a **unit vector** if $\|x\| = 1$. Finally, a subset S of V is **orthonormal** if S is orthogonal and consists entirely of unit vectors*

Note that if $S = \{v_1, v_2, ...\}$, then S is orthonormal if and only if $\langle v_i, v_j \rangle = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta. Also, observe that multiplying vectors by nonzero scalars does not affect their orthogonality and that if x is any nonzero vector, then $(1/\|x\|)x$ is a unit vector. The process of multiplying a nonzero vector by the reciprocal of its length is called **normalizing**.

Recall the inner product space H. We introduce an important orthonormal subset S of H. Now define $S = \{f_n : n \text{ is an integer}\}$ where $f_n(t) = e^{int}$. Clearly S is a subset of H. We can prove that $\langle f_m, f_n \rangle = \delta_{mn}$. See more on textbook (page 335 and 336).

6.2 The Gram-Schmidt Orthogonalization Process and Orthogonal Complements

Definition. *Let V be an inner product space. A subset of V is an **orthonormal basis** for V if it is an ordered basis that is orthonormal.*

Theorem 6.3. *Let V be an inner product space and $S = \{v_1, v_2, ..., v_k\}$ be an orthogonal subset of V consisting of nonzero vectors. If $y \in span(S)$, then*

$$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i$$

Corollary 1. *If, in addition to the hypotheses of Theorem 6.3, S is orthonormal and $y \in span(S)$, then*

$$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$$

Corollary 2. *Let V be an inner product space, and let S be an orthogonal subset of V consisting of nonzero vectors. Then S is linearly independent.*

Theorem 6.4. *Let V be an inner product space and $S = \{w_1, w_2, ..., w_n\}$ be a linearly independent subset of V. Define $S' = \{v_1, v_2, ..., v_n\}$, where $v_1 = w_1$ and*

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j$$

Then S' is an orthogonal set of nonzero vectors such that $span(S') = span(S)$.

The construction of $\{v_1, v_2, ..., v_n\}$ by the use of Theorem 6.4 is called the **Gram–Schmidt process**.

If we continue applying the Gram–Schmidt orthogonalization process to the basis $\{1, x, x^2, ...\}$ for $P(R)$, we obtain an orthogonal basis whose elements are called the *Legendre polynomials*.

Theorem 6.5. *Let V be a nonzero finite-dimensional inner product space. Then V has an orthonormal basis β. Furthermore, if $\beta = \{v_1, v_2, ..., v_n\}$ and $x \in V$, then*

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$$

Corollary. *Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, v_2, ..., v_n\}$. Let T be a linear operator on V, and let $A = [T]_\beta$. Then for any i and j, $A_{ij} = \langle T(v_j), v_i \rangle$.*

Definition. *Let β be an orthonormal subset (possibly infinite) of an inner product space V, and let $x \in V$. We define the **Fourier coefficients** of x relative to β to be the scalars $\langle x, y \rangle$, where $y \in \beta$.*

Definition. *Let S be a nonempty subset of an inner product space V. We define $S^\perp$ (read "S perp") to be the set of all vectors in V that are orthogonal to every vector in S; that is, $S^\perp = \{x \in V : \langle x, y \rangle = 0$ for all $y \in S\}$. The set $S^\perp$ is called the **orthogonal complement** of S.*

Theorem 6.6. *Let W be a finite-dimensional subspace of an inner product space V, and let $y \in V$. Then there exist unique vectors $u \in W$ and $z \in W^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, v_2, ..., v_k\}$ is an orthonormal basis for W, then*

$$u = \sum_{i=1}^{k} \langle y, u_i \rangle v_i$$

Corollary. *In the notation of Theorem 6.6, the vector u is the unique vector in W that is "closest" to y; that is, for any $x \in W$, $\|y - x\| \geq \|y - u\|$, and this inequality is an equality if and only if $x = u$.*

The vector u in the corollary is called the **orthogonal projection** of y on W.

Theorem 6.7. *Suppose that $S = \{v_1, v_2, ..., v_k\}$ is an orthonormal set in an n-dimensional inner product space V. Then*

(a) *S can be extended to an orthonormal basis $\{v_1, v_2, ..., v_k, v_{k+1}, ..., v_n\}$ for V.*

(b) *If $W = span(S)$, then $S_1 = \{v_{k+1}, v_{k+2}, ..., v_n\}$ is an orthonormal basis for $W^\perp$ (using the preceding notation).*

(c) *If W is any subspace of V, then $\dim(V) = \dim(W) + \dim(W^\perp)$.*

6.3   The Adjoint of a Linear Operator

Theorem 6.8. *Let V be a finite-dimensional inner product space over F, and let $g : V \to F$ be a linear transformation. Then there exists a unique vector $y \in V$ such that $g(x) = \langle x, y \rangle$ for all $x \in V$.*

Theorem 6.9. *Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Then there exists a unique function $T^* : V \rightarrow V$ such that $\langle T(x), y \rangle = \langle x, T^*(y) \rangle$ for all $x, y \in V$. Furthermore, T^* is linear.*

The linear operator T^* described in Theorem 6.9 is called the **adjoint** of the operator T. The symbol T^* is read "T star."

Thus T^* is the unique operator on V satisfying $\langle T(x), y \rangle = \langle x, T^*(y) \rangle$ for all $x, y \in V$. Note that we also have

$$\langle x, T(y) \rangle = \overline{\langle T(y), x \rangle} = \overline{\langle y, T^*(x) \rangle} = \langle T^*(x), y \rangle$$

so $\langle x, T(y) \rangle = \langle T^*(x), y \rangle$ for all $x, y \in V$. We may view these equations symbolically as adding a * to T when shifting its position inside the inner product symbol.

Theorem 6.10. *Let V be a finite-dimensional inner product space, and let β be an orthonormal basis for V. If T is a linear operator on V, then*

$$[T^*]_\beta = [T]_\beta^*$$

Corollary. *Let A be an $n \times n$ matrix. Then $L_{A^*} = (L_A)^*$.*

Theorem 6.11. *Let V be an inner product space, and let T and U be linear operators on V. Then*

(a) $(T + U)^* = T^* + U^*$;

(b) $(cT)^* = \bar{c}T^*$ *for any $c \in F$;*

(c) $(TU)^* = U^*T^*$;

(d) $T^{**} = T$;

(e) $I^* = I$.

Corollary. *Let A and B be $n \times n$ matrices. Then*

(a) $(A + B)^* = A^* + B^*$;

(b) $(cA)^* = \bar{c}A^*$ *for all $c \in F$;*

(c) $(AB)^* = B^*A^*$;

(d) $A^{**} = A$;

(e) $I^* = I$.

Least Squares Approximation

Lemma1. *Let $A \in M_{m \times n}(F), x \in F^n$, and $y \in F^m$. Then*

$$\langle Ax, y \rangle_m = \langle x, A^*y \rangle_n$$

Lemma2. *Let $A \in M_{m \times n}(F)$. Then $rank(A^*A) = rank(A)$.*

Corollary. *If A is an $m \times n$ matrix such that $rank(A) = n$, then A^*A is invertible.*

Theorem 6.12. *Let $A \in M_{m \times n}(F)$ and $y \in F_m$. Then there exists $x_0 \in F_n$ such that $(A^*A)x_0 = A^*y$ and $\|Ax_0 - y\| \leq \|Ax - y\|$ for all $x \in F^n$. Furthermore, if $rank(A) = n$, then $x_0 = (A^*A)^{-1}A^*y$.*

Minimal Solutions to Systems of Linear Equations

Theorem 6.13. *Let $A \in M_{m \times n}(F)$ and $b \in F^m$. Suppose that $Ax = b$ is consistent. Then the following statements are true.*
  (a) There exists exactly one minimal solution s of $Ax = b$, and $s \in R(L_{A^*})$.
  (b) The vector s is the only solution to $Ax = b$ that lies in $R(L_{A^*})$; that is, if u satisfies $(AA^*)u = b$, then $s = A^*u$.

Definition. *Let $T : V \to W$ be a linear transformation, where V and W are finite-dimensional inner product spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively. A function $T^* : W \langle V \rangle$ is called an adjoint of T if $\langle T(x), y \rangle_2 = \langle x, T^*(y) \rangle_1$ for all $x \in V$ and $y \in W$.*

6.4   Normal and Self-Adjoint Operators

Lemma. *Let T be a linear operator on a finite-dimensional inner product space V. If T has an eigenvector, then so does T^*.*

Theorem 6.14 (Schur)**.** *Let T be a linear operator on a finite-dimensional inner product space V. Suppose that the characteristic polynomial of T splits. Then there exists an orthonormal basis β for V such that the matrix $[T]_\beta$ is upper triangular.*

Definitions. *Let V be an inner product space, and let T be a linear operator on V. We say that T is **normal** if $TT^* = T^*T$. An $n \times n$ real or complex matrix A is **normal** if $AA^* = A^*A$.*

Theorem 6.15. *Let V be an inner product space, and let T be a normal operator on V. Then the following statements are true.*

  (a) $\|T(x)\| = \|T^*(x)\|$ for all $x \in V$.

(b) $T - cI$ is normal for every $c \in F$.

(c) If x is an eigenvector of T, then x is also an eigenvector of T^*. In fact, if $T(x) = \lambda x$, then $T^*(x) = \overline{\lambda} x$.

(d) If λ_1 and λ_2 are distinct eigenvalues of T with corresponding eigenvectors x_1 and x_2, then x_1 and x_2 are orthogonal.

Theorem 6.16. *Let T be a linear operator on a finite-dimensional complex inner product space V. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.*

Definitions. *Let T be a linear operator on an inner product space V. We say that T is **self-adjoint (Hermitian)** if $T = T^*$. An $n \times n$ real or complex matrix A is **self-adjoint (Hermitian)** if $A = A^*$.*

Lemma. *Let T be a self-adjoint operator on a finite-dimensional inner product space V. Then*
(a) Every eigenvalue of T is real.
(b) Suppose that V is a real inner product space. Then the characteristic polynomial of T splits.

Theorem 6.17. *Let T be a linear operator on a finite-dimensional real inner product space V. Then T is self-adjoint if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.*

Definitions. *A linear operator T on a finite-dimensional inner product space is called **positive definite** [**positive semidefinite**] if T is self-adjoint and $\langle T(x), x \rangle > 0 [\langle T(x), x \rangle \geq 0]$ for all $x \neq 0$.*
*An $n \times n$ matrix A with entries from R or C is called **positive definite** [**positive semidefinite**] if L_A is positive definite [positive semidefinite].*

6.5 Unitary and Orthogonal Operators and Their Matrices

Definitions. *Let T be a linear operator on a finite-dimensional inner product space V (over F). If $\|T(x)\| = \|x\|$ for all $x \in V$, we call T a **unitary operator** if $F = C$ and an **orthogonal operator** if $F = R$.*

It should be noted that, in the infinite-dimensional case, an operator satisfying the preceding norm requirement is generally called an **isometry**. If, in addition, the operator is onto (the condition guarantees one-to-one), then the operator is called a **unitary** or **orthogonal operator**.

Theorem 6.18. *Let T be a linear operator on a finite-dimensional inner product space V. Then the following statements are equivalent.*

(a) $TT^* = T^*T = I$.

(b) $\langle T(x), T(y) \rangle = \langle x, y \rangle$ for all $x, y \in V$.

(c) If β is an orthonormal basis for V, then $T(\beta)$ is an orthonormal basis for V.

(d) There exists an orthonormal basis β for V such that $T(\beta)$ is an orthonormal basis for V.

(e) $\|T(x)\| = \|x\|$ for all $x \in V$.

Lemma. *Let U be a self-adjoint operator on a finite-dimensional inner product space V. If $\langle x, U(x) \rangle = 0$ for all $x \in V$, then $U = T_0$.*

Corollary 1. *Let T be a linear operator on a finite-dimensional real inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is both self-adjoint and orthogonal.*

Corollary 2. *Let T be a linear operator on a finite-dimensional complex inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is unitary.*

Definition. *Let L be a one-dimensional subspace of R^2. We may view L as a line in the plane through the origin. A linear operator T on R^2 is called a **reflection of** R^2 **about** L if $T(x) = x$ for all $x \in L$ and $T(x) = -x$ for all $x \in L^\perp$.*

Definitions. *A square matrix A is called an an **orthogonal matrix** if $A^t A = AA^t = I$ and **unitary** if $A^* A = AA^* = I$.*

Since for a real matrix A we have $A^* = A^t$, a real unitary matrix is also orthogonal. In this case, we call A **orthogonal** rather than unitary.

Note that the condition $AA^* = I$ is equivalent to the statement that the rows of A form an orthonormal basis for F^n. A similar remark can be made about the columns of A and the condition $A^* A = I$.

It also follows from the definition above and from Theorem 6.10 (p. 359) that a linear operator T on an inner product space V is unitary [orthogonal] if and only if $[T]_\beta$ is unitary [orthogonal] for some orthonormal basis β for V.

We know that, for a complex normal [real symmetric] matrix A, there exists an orthonormal basis β for F^n consisting of eigenvectors of A. Hence A is similar to a diagonal matrix D. By the corollary to Theorem 2.23 (p. 115), the matrix Q whose columns are the vectors in β is such that $D = Q^{-1}AQ$. But since the columns of Q are an orthonormal basis for F^n, it follows that Q is unitary [orthogonal]. In this case, we say that A is **unitarily equivalent** [**orthogonally equivalent**] to D. It is easily seen (see Exercise 18) that this relation is an equivalence relation on $M_{n\times n}(C)$ $[M_{n\times n}(R)]$. More generally, A and B are unitarily equivalent [orthogonally equivalent] if and only if there exists a unitary [orthogonal] matrix P such that $A = P^*BP$.

Theorem 6.19. *Let A be a complex $n \times n$ matrix. Then A is normal if and only if A is unitarily equivalent to a diagonal matrix.*

Theorem 6.20. *Let A be a real $n \times n$ matrix. Then A is symmetric if and only if A is orthogonally equivalent to a real diagonal matrix.*

Theorem 6.21 (Schur). *Let $A \in M_{n \times n}(F)$ be a matrix whose characteristic polynomial splits over F.*

(a) *If $F = C$, then A is unitarily equivalent to a complex upper triangular matrix.*

(b) *If $F = R$, then A is orthogonally equivalent to a real upper triangular matrix.*

Rigid Motions*

Theorem 6.22.

Orthogonal Operators on R^2

Theorem 6.23.

Conic Sections .

6.6 Orthogonal Projections and the Spectral Theorem

Definition. *Let V be an inner product space, and let T: V $\to$ V be a projection. We say that T is an **orthogonal projection** if $R(T)^\perp = N(T)$ and $N(T)^\perp = R(T)$.*

In the notation of Theorem 6.6, we can define a function $T : V \to V$ by T(y) = u. We call T the **orthogonal projection** of V on W.

Theorem 6.24. *Let V be an inner product space, and let T be a linear operator on V. Then T is an orthogonal projection if and only if T has an adjoint T^* and $T^2 = T = T^*$.*

Theorem 6.25 (The Spectral Theorem). *Suppose that T is a linear operator on a finite-dimensional inner product space V over F with the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Assume that T is normal if F = C and that T is self-adjoint if F = R. For each i ($1 \leq i \leq k$), let W_i be the eigenspace of T corresponding to the eigenvalue λ_i, and let T_i be the orthogonal projection of V on W_i. Then the following statements are true.*

(a) $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

(b) *If W_i' denotes the direct sum of the subspaces W_j for $j \neq i$, then $W_i^\perp = W_i'$.*

(c) $T_i T_j = \delta_{ij} T_i$ *for $1 \leq i, j \leq k$.*

(d) $I = T_1 + T_2 + \cdots + T_k$.

(e) $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$.

The set $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ of eigenvalues of T is called the **spectrum** of T, the sum $I = T_1 + T_2 + \cdots + T_k$ in (d) is called the **resolution of the identity operator** induced by T, and the sum $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ in (e) is called the **spectral decomposition** of T. The spectral decomposition of T is unique up to the order of its eigenvalues.

If $\lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ is the spectral decomposition of T, then it follows (from Exercise 7) that $g(T) = g(\lambda_1)T_1 + g(\lambda_2)T_2 + \cdots + g(\lambda_k)T_k$ for any polynomial g.

Corollary 1. *If $F = C$, then T is normal if and only if $T^* = g(T)$ for some polynomial g.*

Corollary 2. *If $F = C$, then T is unitary if and only if T is normal and $|\lambda| = 1$ for every eigenvalue λ of T.*

Corollary 3. *If $F = C$ and T is normal, then T is self-adjoint if and only if every eigenvalue of T is real.*

Corollary 4. *Let T be as in the spectral theorem with spectral decomposition $\lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$. Then each T_j is a polynomial in T.*

6.7 The Singular Value Decomposition and the Pseudoinverse*

Theorem 6.26.

Theorem 6.27.

Theorem 6.28.

Theorem 6.29.

Theorem 6.30.

6.8 Bilinear and Quadratic Forms*

Bilinear Forms

Definition. *Let V be a vector space over a field F. A function H from the set $V \times V$ of ordered pairs of vectors to F is called a **bilinear form** on V if H is linear in each variable when the other variable is held fixed; that is, H is a bilinear form on V if*

(a) $H(ax_1 + x_2, y) = aH(x_1, y) + H(x_2, y)$ for all $x_1, x_2, y \in V$ and $a \in F$

(b) $H(x, ay_1 + y_2) = aH(x, y_1) + H(x, y_2)$ for all $x, y_1, y_2 \in V$ and $a \in F$.

We denote the set of all bilinear forms on V by $\mathcal{B}(V)$. Observe that an inner product on a vector space is a bilinear form if the underlying field is real, but not if the underlying field is complex.

For any bilinear form H on a vector space V over a field F, the following properties hold.

1. If, for any $x \in V$, the functions $L_x, R_x : V \to F$ are defined by

$$L_x(y) = H(x, y) \text{ and } R_x(y) = H(y, x) \text{ for all } y \in V$$

then L_x and R_x are linear.

2. $H(0, x) = H(x, 0) = 0$ for all $x \in V$.

3. For all $x, y, z, w \in V$,

$$H(x + y, z + w) = H(x, z) + H(x, w) + H(y, z) + H(y, w)$$

4. If $J : V \times V \to F$ is defined by $J(x, y) = H(y, x)$, then J is a bilinear form.

Definitions. *Let V be a vector space, let H_1 and H_2 be bilinear forms on V, and let a be a scalar. We define the **sum** $H_1 + H_2$ and the **scalar product** aH_1 by the equations*

$$(H_1 + H_2)(x, y) = H_1(x, y) + H_2(x, y)$$

and

$$(aH_1)(x, y) = a(H_1(x, y)) \text{ for all } x, y \in V$$

Theorem 6.31. *For any vector space V, the sum of two bilinear forms and the product of a scalar and a bilinear form on V are again bilinear forms on V. Furthermore, $\mathcal{B}(V)$ is a vector space with respect to these operations.*

Let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for an n-dimensional vector space V, and let $H \in \mathcal{B}(V)$. We can associate with H an $n \times n$ matrix A whose entry in row i and column j is defined by

$$A_{ij} = H(v_i, v_j) \text{ for } i, j = 1, 2, \ldots, n$$

Definition. *The matrix A above is called the **matrix representation** of H with respect to the ordered basis β and is denoted by $\psi_\beta(H)$.*

Theorem 6.32. *For any n-dimensional vector space V over F and any ordered basis β for V, $\psi_\beta : \mathcal{B}(V) \to M_{n \times n}(F)$ is an isomorphism.*

Corollary 1. *For any n-dimensional vector space V, $\mathcal{B}(V)$ has dimension n^2.*

Corollary 2. *Let V be an n-dimensional vector space over F with ordered basis β. If $H \in \mathcal{B}(V)$ and $A \in M_{n \times n}(F)$, then $\psi_\beta(H) = A$ if and only if $H(x, y) = [\phi_\beta(x)]^t A [\phi_\beta(y)]$ for all $x, y \in V$.*

Corollary 3. *Let F be a field, n a positive integer, and β be the standard ordered basis for F^n. Then for any $H \in \mathcal{B}(F^n)$, there exists a unique matrix $A \in M_{n \times n}(F)$, namely, $A = \psi_\beta(H)$, such that*

$$H(x, y) = x^t A y \text{ for all } x, y \in F^n.$$

Definition. *Let $A, B \in M_{n \times n}(F)$. Then B is said to be **congruent** to A if there exists an invertible matrix $Q \in M_{n \times n}(F)$ such that $B = Q^t A Q$.*

Observe that the relation of congruence is an equivalence relation.

Theorem 6.33. *Let V be a finite-dimensional vector space with ordered bases $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_n\}$, and let Q be the change of coordinate matrix changing γ-coordinates into β-coordinates. Then, for any $H \in \mathcal{B}(V)$, we have $\psi_\gamma(H) = Q^t \psi_\beta(H) Q$. Therefore $\psi_\gamma(H)$ is congruent to $\psi_\beta(H)$.*

Corollary. *Let V be an n-dimensional vector space with ordered basis β, and let H be a bilinear form on V. For any $n \times n$ matrix B, if B is congruent to $\psi_\beta(H)$, then there exists an ordered basis γ for V such that $\psi_\gamma(H) = B$. Furthermore, if $B = Q^t \psi_\beta(H) Q$ for some invertible matrix Q, then Q changes γ-coordinates into β-coordinates.*

Symmetric Bilinear Forms

Definition. *A bilinear form H on a vector space V is **symmetric** if $H(x, y) = H(y, x)$ for all $x, y \in V$.*

Theorem 6.34. *Let H be a bilinear form on a finite-dimensional vector space V, and let β be an ordered basis for V. Then H is symmetric if and only if $\psi_\beta(H)$ is symmetric.*

Definition. *A bilinear form H on a finite-dimensional vector space V is called **diagonalizable** if there is an ordered basis β for V such that $\psi_\beta(H)$ is a diagonal matrix.*

Corollary. *Let H be a diagonalizable bilinear form on a finite-dimensional vector space V. Then H is symmetric.*

Lemma. *Let H be a nonzero symmetric bilinear form on a vector space V over a field F not of characteristic two. Then there is a vector x in V such that $H(x, x) \neq 0$.*

Theorem 6.35. *Let V be a finite-dimensional vector space over a field F not of characteristic two. Then every symmetric bilinear form on V is diagonalizable.*

Corollary. *Let F be a field that is not of characteristic two. If $A \in M_{n \times n}(F)$ is a symmetric matrix, then A is congruent to a diagonal matrix.*

Diagonalization of Symmetric Matrices .

Quadratic Forms

Definition. *Let V be a vector space over F. A function $K : V \to F$ is called a* ***quadratic form*** *if there exists a symmetric bilinear form $H \in \mathcal{B}(V)$ such that*

$$K(x) = H(x, x) \text{ for all } x \in V$$

If the field F is not of characteristic two, there is a one-to-one correspondence between symmetric bilinear forms and quadratic forms given by

$$H(x, y) = \frac{1}{2}[K(x + y) - K(x) - K(y)]$$

Quadratic Forms Over the Field R

Theorem 6.36. *Let V be a finite-dimensional real inner product space, and let H be a symmetric bilinear form on V. Then there exists an orthonormal basis β for V such that $\psi_\beta(H)$ is a diagonal matrix.*

Corollary. *Let K be a quadratic form on a finite-dimensional real inner product space V. There exists an orthonormal basis $\beta = \{v_1, v_2, \ldots, v_n\}$ for V and scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ (not necessarily distinct) such that if $x \in V$ and*

$$x = \sum_{i=1}^{n} s_i v_i, \ \ s_i \in R$$

then

$$K(x) = \sum_{i=1}^{n} \lambda_i s_i^2$$

In fact, if H is the symmetric bilinear form determined by K, then β can be chosen to be any orthonormal basis for V such that $\psi_\beta(H)$ is a diagonal matrix.

The Second Derivative Test for Functions of Several Variables .

Sylvester's Law of Inertia .

7 Canonical Forms

7.1 The Jordan Canonical Form I

In this section, we extend the definition of eigenspace to *generalized eigenspace.* From these subspaces, we select ordered bases whose union is an ordered basis β for V such that

$$[T]_\beta = \begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & A_k \end{pmatrix}$$

43

, where each O is a zero matrix, and each A_i is a square matrix of the form (λ) or

$$
\begin{pmatrix}
\lambda & 1 & 0 & \cdots & 0 & 0 \\
0 & \lambda & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \lambda & 1 \\
0 & 0 & 0 & \cdots & 0 & \lambda
\end{pmatrix}
$$

for some eigenvalue λ of T. Such a matrix A_i is called a **Jordan block** corresponding to λ, and the matrix $[T]_\beta$ is called a **Jordan canonical form** of T. We also say that the ordered basis β is a **Jordan canonical basis** for T. Observe that each Jordan block A_i is "almost" a diagonal matrix—in fact, $[T]_\beta$ is a diagonal matrix if and only if each A_i is of the form (λ).

Definition. *Let T be a linear operator on a vector space V, and let λ be a scalar. A nonzero vector x in V is called a **generalized eigenvector of** T **corresponding to** λ if $(T-\lambda I)^p(x) = 0$ for some positive integer p.*

Definition. *Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. The **generalized eigenspace of** T **corresponding to** λ, denoted K_λ, is the subset of V defined by*

$$
K_\lambda = \{x \in V : (T - \lambda I)^p(x) = 0 \text{ for some positive integer } p\}
$$

Theorem 7.1. *Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. Then*

(a) *K_λ is a T-invariant subspace of V containing E_λ (the eigenspace of T corresponding to λ).*

(b) *For any scalar $\mu \neq \lambda$, the restriction of $T-\mu I$ to K_λ is one-to-one.*

Theorem 7.2. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Suppose that λ is an eigenvalue of T with multiplicity m. Then*

(a) *$dim(K_\lambda) \leq m$.*

(b) *$K_\lambda = N((T - \lambda I)^m)$.*

Theorem 7.3. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then, for every $x \in V$, there exist vectors $v_i \in K_{\lambda_i}$, $1 \leq i \leq k$, such that*

$$
x = v_1 + v_2 + \cdots + v_k
$$

Theorem 7.4. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T with corresponding multiplicities $m_1, m_2, \ldots, m_k$. For $1 \leq i \leq k$, let β_i be an ordered basis for K_{λ_i}. Then the following statements are true.*

(a) *$\beta_i \cap \beta_j = \varnothing$ for $i \neq j$.*

(b) *$\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is an ordered basis for V.*

(c) *$dim(K_{\lambda_i}) = m_i$ for all i.*

Corollary. *Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Then T is diagonalizable if and only if $E_\lambda = K_\lambda$ for every eigenvalue λ of T.*

Definitions. *Let T be a linear operator on a vector space V, and let x be a generalized eigenvector of T corresponding to the eigenvalue λ. Suppose that p is the smallest positive integer for which $(T-\lambda I)^p(x) = 0$. Then the ordered set*

$$\{(T-\lambda I)^{p-1}(x), (T-\lambda I)^{p-2}(x), \lambda, (T-\lambda I)(x), x\}$$

*is called a **cycle of generalized eigenvectors** of T corresponding to λ. The vectors $(T-\lambda I)^{p-1}(x)$ and x are called the **initial vector** and the **end vector** of the cycle, respectively. We say that the **length** of the cycle is p.*

Theorem 7.5. *Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits, and suppose that β is a basis for V such that β is a disjoint union of cycles of generalized eigenvectors of T. Then the following statements are true.*

(a) *For each cycle γ of generalized eigenvectors contained in β, $W = span(\gamma)$ is T-invariant, and $[T_W]_\gamma$ is a Jordan block.*

(b) *β is a Jordan canonical basis for V.*

Theorem 7.6. *Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. Suppose that $\gamma_1, \gamma_2, \ldots, \gamma_q$ are cycles of generalized eigenvectors of T corresponding to λ such that the initial vectors of the γ_i's are distinct and form a linearly independent set. Then the γ_i's are disjoint, and their union $\gamma = \bigcup_{i=1}^{q} \gamma_i$ is linearly independent.*

Corollary. *Every cycle of generalized eigenvectors of a linear operator is linearly independent.*

Theorem 7.7. *Let T be a linear operator on a finite-dimensional vector space V, and let λ be an eigenvalue of T. Then K_λ has an ordered basis consisting of a union of disjoint cycles of generalized eigenvectors corresponding to λ.*

Corollary 1. *Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits. Then T has a Jordan canonical form.*

Definition. *Let $A \in M_{n \times n}(F)$ be such that the characteristic polynomial of A (and hence of L_A) splits. Then the **Jordan canonical form** of A is defined to be the Jordan canonical form of the linear operator L_A on F^n.*

Corollary 2. *Let A be an $n \times n$ matrix whose characteristic polynomial splits. Then A has a Jordan canonical form J, and A is similar to J.*

Theorem 7.8. *Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits. Then V is the direct sum of the generalized eigenspaces of T.*

7.2   The Jordan Canonical Form II

To help us visualize each of the matrices A_i and ordered bases β_i, we use an array of dots called a **dot diagram** of T_i, where T_i is the restriction of T to K_{λ_i}. (See more in the textbook)

 To facilitate our arguments, we fix a basis β_i for K_{λ_i} so that β_i is a disjoint union of n_i cycles of generalized eigenvectors with lengths $p_1 \geq p_2 \geq \cdots \geq p_{n_i}$.

Theorem 7.9. *For any positive integer r, the vectors in β_i that are associated with the dots in the first r rows of the dot diagram of T_i constitute a basis for $N((T-\lambda_i I)^r)$. Hence the number of dots in the first r rows of the dot diagram equals $nullity((T-\lambda_i I)^r)$.*

Corollary. *The dimension of E_{λ_i} is n_i. Hence in a Jordan canonical form of T, the number of Jordan blocks corresponding to λ_i equals the dimension of E_{λ_i}.*

Theorem 7.10. *Let r_j denote the number of dots in the jth row of the dot diagram of T_i, the restriction of T to K_{λ_i}. Then the following statements are true.*

 (a) $r_1 = dim(V) - rank(T - \lambda_i I)$.

 (b) $r_j = rank((T - \lambda_i I)^{j-1}) - rank((T - \lambda_i I)^j)$ if $j > 1$.

 Theorem 7.10 shows that the dot diagram of T_i is completely determined by T and λ_i.

Corollary. *For any eigenvalue λ_i of T, the dot diagram of T_i is unique. Thus, subject to the convention that the cycles of generalized eigenvectors for the bases of each generalized eigenspace are listed in order of decreasing length, the Jordan canonical form of a linear operator or a matrix is unique up to the ordering of the eigenvalues.*

 We do not attempt, however, to develop a general algorithm for computing Jordan canonical bases, although one could be devised by following the steps in the proof of the existence of such a basis (Theorem 7.7 p. 490).

Theorem 7.11. *Let A and B be $n \times n$ matrices, each having Jordan canonical forms computed according to the conventions of this section. Then A and B are similar if and only if they have (up to an ordering of their eigenvalues) the same Jordan canonical form.*

The reader should observe that any diagonal matrix is a Jordan canonical form. Thus a linear operator T on a finite-dimensional vector space V is diagonalizable if and only if its Jordan canonical form is a diagonal matrix. Hence T is diagonalizable if and only if the Jordan canonical basis for T consists of eigenvectors of T. Similar statements can be made about matrices.