Linear Algebra Notes

Wei-Hsiang Sun

Feb. 2021

Definitions and Theorems of Linear Algebra from Linear Algebra S. Friedberg A. Insel L. Spence Fourth Edition

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.

Parallel vectors x and y are parallel if $y = tx$ for some nonzero real number t

1.2 Vector Spaces

Definition. A **vector space** (or **linear space**) V over a field F consists of a set on which two operations (called **addition** and **scaler 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 element $a,\ b$ in F and each element x in V, $(a + b)x = ax + by$.

Different types of vectors.

- **N-tuple** an object of the form $(a_1,\ a_2, \cdots\ a_n)$, where entries $a_1,\ a_2, \cdots\ a_n$ are elements of a field F. $a_1,\ a_2, \cdots\ a_n$ are called entries or the components of the n-tuple. An n-tuple can also be denoted by F^n, which may be written as a column vector.

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

- **Matrix** An $m \times n$ matrix with entries from a field F is a rectangular array of the form

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1_n} \\ a_{21} & a_{22} & \dots & a_{2_n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

the set of all $m \times n$ matrices with entries from the field F can also be denoted by $\boldsymbol{M}_{m\times n}(F)$, some properties of the matrix are:
diagonal entries, row(F^m), column(F^n), square matrix, zero matrix, two matrix equal condition

- **Functions** Let S be any non-empty set and F be any field, and let $\mathcal{F}(S,F)$ denote the set of all functions from S to F. Two functions f and g in $\mathcal{F}(S,F)$ are called equal if $f(s) = g(s)$ for each $s \in S$. Let $c \in F$, the addition and multiplication are defined,

$$(f+g)(s) = f(s) + g(s) \quad \text{and} \quad (cf)(s) = c[f(s)]$$

The set $\mathcal{F}(S,F)$ is a vector space

- **Polynomial** a polynomial with coefficients from a field F is an expression of the form

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

some properties of polynomials:
coefficient, zero polynomial and its degree (-1), degree of a polynomial (largest exponent that appears in the representation)

- **Sequence** Let F be any field. A sequence in F is a function σ from the positive integers into F. A sequence s.t. $\sigma(n) = a_n$ for $n = 1, 2, \cdots$ is denoted $\{a_n\}$. Let V be the set consist of all sequences in F, and $\{a_n\}$, $\{b_n\}$ are in V, $t \in F$. Then the addition and multiplication are defined as follows:

$$\{a_n\} + \{b_n\} = \{a_n + b_n\} \quad \text{and} \quad t\{a_n\} = \{ta_n\}$$

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 1.1.1. The vector 0 described in (VS 3) is unique.

Corollary 1.1.2. The vector y describe in (VS 4) is unique.

Note: y in (VS 4) is called the additive inverse 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.

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 definitions 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 $x \in W$

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 $(A^t)_{ij} = A_{ji}$.

Symmetric Matrix is a matrix A such that $A^t = A$.

Diagonal Matrix A $n \times n$ matrix M is a diagonal matrix if $M_{ij} = 0$ whenever $i \neq j$.

Trace of an $n \times n$ matrix M, denoted $\mathrm{tr}(M) = M_{11} + M_{22} + \cdots + M_{nn}$.

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

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 $u_1, u_2, \cdots, u_n$ in S and scalars $a_1, a_2, \cdots, a_n$ in F such that $v = a_1u_1 + a_2u_2 + \cdots + a_nu_n$. In this case we also say that v is a linear combination of $u_1, u_2, \cdots, u_n$ and call $a_1, a_2, \cdots, a_n$ the **coefficients** of the linear combination.

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

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 vector space V is called **linearly independent** if there exist a finite number of distinct vectors $u_1, u_2, \cdots, u_n$ in S and scalers $a_1, a_2, \cdots, a_n$, not all zero, such that

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0$$

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

Trivial Representation For any vectors $u_1, u_2, \cdots, u_n$, we have $a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0$ if $a_1 = a_2 = \cdots = a_n = 0$. We call this the **trivial representation** of 0 as a linear combination of $u_1, u_2, \cdots, u_n$

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.

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 1.6.1. 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 $v \in \text{span}(S)$.

1.6 Bases 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 vectors of β form a basis for V.

Theorem 1.8. Let V be a vector space and $\beta = \{u_1, u_2, \cdots, u_n\}$ 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

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n$$

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

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

Definition. 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 1.10.2. 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 $\dim(\mathsf{W}) \leq \dim(\mathsf{V})$. Moreover, if $\dim(\mathsf{W}) = \dim(\mathsf{V})$, then $\mathsf{W} = \mathsf{V}$

1.7 Maximal Linearly Independent Subsets *

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

Power set Let $\mathcal{F}$ be a family of all subsets of a nonempty set S, the $\mathcal{F}$ is a **power set** of S.

Definition. A collection of sets $\mathcal{C}$ is called a **chain** (or **nest** or **tower**) if for each pair of sets A and B in $\mathcal{C}$, either $A \subseteq B$ or $B \subseteq A$

Maximal Principle Let $\mathcal{F}$ be a family of sets. If, for each chain $\mathcal{C} \subseteq \mathcal{F}$ there exists a member of $\mathcal{F}$ that contains each member of $\mathcal{C}$, then $\mathcal{F}$ 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.

Theorem 1.12. Let V be a vector space and S a subset that generates V If β is 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 1.13.1. Every vector space has a basis.

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 $\mathsf{T} : \mathsf{W} \to \mathsf{V}$ a **linear transformation from** V **to** W if, for all $x, y \in \mathsf{V}$ and $c \in F$, we have

(a) $\mathsf{T}(x + y) = \mathsf{T}(x) + \mathsf{T}(y)$

(b) $\mathsf{T}(cx) = c\mathsf{T}(x)$

Note: T can be simply called **linear**

For any angle θ, define $\mathsf{T}_\theta : \mathsf{R}^2 \to \mathsf{R}^2$ by the rule: $\mathsf{T}_\theta(a_1, a_2)$ is the vector obtained b rotating (a_1, a_2) counterclockwise by θ if $(a_1, a_2) \neq (0, 0)$, and $\mathsf{T}_\theta(0, 0) = (0, 0)$. Then $\mathsf{T}_\theta : \mathsf{R}^2 \to \mathsf{R}^2$ is a linear transformation that is called the **rotation by** θ.

Define $\mathsf{T}_\theta : \mathsf{R}^2 \to \mathsf{R}^2$ by $\mathsf{T}(a_1, a_2) = (a_1, -a_2)$. T is called the **reflection about the** x**-axis**.

Define $\mathsf{T}_\theta : \mathsf{R}^2 \to \mathsf{R}^2$ by $\mathsf{T}(a_1, a_2) = (a_1, 0)$. T is called the **projection on the** x**-axis**.

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

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

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

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

$$\mathsf{R}(\mathsf{T}) = \text{span}(\mathsf{T}(\beta)) = \text{span}(\{\mathsf{T}(v_1), \mathsf{T}(v_2, \ldots, \mathsf{T}(v_n\}).$$

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

Theorem 2.3 (**Dimension Theorem**)**.** Let V and W be vector spaces and $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ be linear. If V is finite-dimensional, then

$$\text{nullity}(\mathsf{T}) + \text{rank}(\mathsf{T}) = \dim(\mathsf{V})$$

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

Theorem 2.5. Let V and W be vector spaces of equal(finite) dimension, and let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ be linear. Then the following are equivalent.

(a) T is one-to-one.

(b) T is onto

(c) $\text{rank}(\mathsf{T}) = \dim(\mathsf{V})$

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 $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ such that $\mathsf{T}(v_i) = w_i$ for $i = 1, 2, \ldots, n$.

Corollary 2.6.1. Let V and W be vector spaces, and suppose that V has a finite basis $\{v_1, v_2, \ldots, v_n\}$. If $\mathsf{U}, \mathsf{T} : \mathsf{V} \to \mathsf{W}$ are linear and $\mathsf{U}(v_i) = \mathsf{T}(v_i)$ for $i = 1, 2, \ldots, n$, then $\mathsf{U} = \mathsf{T}$

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 a 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 $\mathsf{P}_n(F)$, we call $\{1, x, x^2, \ldots, x^n\}$ the **standard ordered basis** for $\mathsf{P}_n(F)$

Definition. Let $\beta = \{u_1, u_2, \ldots, u_n\}$ be an ordered basis for finite-dimensional vector space V. For $x \in \mathsf{V}$, 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

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

Notation: Suppose that V and W are finite-dimensional vector spaces with ordered bases $\beta = \{v_1, v_2, \ldots, v_n\}$, $\gamma = \{w_1, w_2, \ldots, w_n\}$, respectively. Let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ 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

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

Definition. 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 $A = [\mathsf{T}]_\beta^\gamma$ and write $A = [\mathsf{T}]_\beta^\gamma$. If $\mathsf{V} = \mathsf{W}$ and $\beta = \gamma$, then we write $A = [\mathsf{T}]_\beta$

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

Theorem 2.7. Let V and W be vector spaces over a field F, and let $\mathsf{T}, \mathsf{U} : \mathsf{V} \to \mathsf{W}$ be linear.

(a) For all $a \in F$, $a\mathsf{T} + \mathsf{U}$ is linear

(b) Using the operations of addition and scaler multiplication in the preceding definition, the collection of all linear transformations from V to W is a vector space over F

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

Theorem 2.8. Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively, and let $\mathsf{T}, \mathsf{U} : \mathsf{V} \to \mathsf{W}$ be linear transformations. Then

(a) $[\mathsf{T} + \mathsf{U}]_\beta^\gamma = [\mathsf{T}]_\beta^\gamma + [\mathsf{U}]_\beta^\gamma$ and

(b) $[a\mathsf{T}]_\beta^\gamma = a[\mathsf{T}]_\beta^\gamma$ for all scalars a.

2.3 Composition of Linear Transformations and Matrix Multiplication

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

Theorem 2.10. Let V be a vector space. Let $\mathsf{T}, \mathsf{U}_1, \mathsf{U}_2 \in \mathcal{L}(\mathsf{V})$. Then

(a) $\mathsf{T}(\mathsf{U}_1 + \mathsf{U}_2) = \mathsf{TU}_1 + \mathsf{TU}_2$ and $(\mathsf{U}_1 + \mathsf{U}_2)\mathsf{T} = \mathsf{U}_1\mathsf{T} + \mathsf{U}_2\mathsf{T}$

(b) $\mathsf{T}(\mathsf{U}_1\mathsf{U}_2) = (\mathsf{TU}_1)\mathsf{U}_2$

(c) $\mathsf{TI} = \mathsf{IT} = \mathsf{T}$

(d) $a(\mathsf{U}_1\mathsf{U}_2) = (a\mathsf{U}_1 U_2) = \mathsf{U}_1(a\mathsf{U}_2)$ 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

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

Theorem 2.11. Let V, W, and Z be finite-dimensional vector spaces with ordered basis $\alpha, \beta$and γ, respectively. Let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ and $\mathsf{U} : \mathsf{W} \to \mathsf{Z}$ be linear transformations. Then

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

Corollary 2.11.1. Let V be a finite-dimensional vector space with an ordered basis β. Let $\mathsf{T}, \mathsf{U} \in \mathcal{L}(\mathsf{V})$. Then $[\mathsf{UT}]_\beta = [\mathsf{U}]_\beta [\mathsf{T}]_\beta$

Definition. We define the **Kronecker delta** δ_{ij} by $\delta_{ij=1}$ if $i = j$ and $\delta_{ij=0}$ if $i \neq j$. 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) $a(AB) = a(A)B = A(aB)$ for any scalar a.

(c) $I_m A = A = AI_n$

(d) If V is a finite-dimensional vector space with an ordered basis β, then $[\mathsf{I_V}]_\beta = I_n$

Corollary 2.12.1. Let A be an $m \times n$ matrix, $B_1, B_2, \ldots, B_k$ be $n \times p$ matrices, $C_1, C)2, \ldots C_k$ be $q \times m$ matrices, and $a_1, 1_2, \ldots, a_k$ be scalers. Then

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

and

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

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

(a) $u_j = Av_j$

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

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

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

Definition. Let A be an $m \times n$ matrix with entries from a field F. We denote by L_A the mapping $\mathsf{L}_A : \mathsf{F}^n \to \mathsf{F}^m$ defined by $\mathsf{L}_A(x) = Ax$(the matrix product of A and x) for each column vector $x \in \mathsf{F}^n$. 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 $\mathsf{L}_A : \mathsf{F}^n \to \mathsf{F}^m$ 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 $\mathsf{F}^n \to \mathsf{F}^m$, respectively, then we have the following properties.

(a) $[\mathsf{L}_A]_\beta^\gamma = A$.

(b) $\mathsf{L}_A = \mathsf{L}_B$ if and only if $A = B$.

(c) $\mathsf{L}_{A+B} = \mathsf{L}_A + \mathsf{L}_B$ and $\mathsf{L}_{aA} = a\mathsf{L}_A$ for all $a \in F$.

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

(e) If E is an $n \times p$ matrix, then $\mathsf{L}_{AE} = \mathsf{L}_A \mathsf{L}_E$

(f) If $m = n$, then $\mathsf{L}_{I_n} = \mathsf{I}_{\mathsf{F}^n}$

Theorem 2.16. Let A, B and C be matrices such that $A(BC)$ is defientd. 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 $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ be linear. A function $\mathsf{U} : \mathsf{W} \to \mathsf{V}$ is said to be an **inverse** of T if $\mathsf{TU} = \mathsf{I}_\mathsf{W}$ and $\mathsf{UT} = \mathsf{I}_\mathsf{V}$. 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}

Facts that hold for invertible functions.

1. $(\mathsf{TU})^{-1} = \mathsf{U}^{-1}\mathsf{T}^{-1}$
2. $(\mathsf{T}^{-1})^{-1} = \mathsf{T}$; in particular, T^{-1} is invertible
3. Let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ be a linear transformation, where V and W are finite dimensional spaces of equal dimension. Then T is invertible if and only if $\mathsf{rank}(\mathsf{T}) = \dim(\mathsf{V})$

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

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

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, $\dim(\mathsf{V}) = \dim(\mathsf{W})$.

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

Corollary 2.18.1. Let V be a finite-dimensional vector space with an ordered basis β, and let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ be linear. Then T is invertible if and only if $[\mathsf{T}]_\beta$ is invertible. Furthermore, $[\mathsf{T}^{-1}]_\beta = ([\mathsf{T}]_\beta)^{-1}$

Corollary 2.18.2. Let A be an $n \times n$ matrix. Then A is invertible if and only if L_A is invertible. Furthermore, $(\mathsf{L}_A)^{-1} = \mathsf{L}_{A^{-1}}$

Definition. Let V and $\mathbb{W}$ be vector spaces. We say that V is **isomorphic** to W if there exists a linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ that is invertible. Such 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 $\dim(\mathsf{V}) = \dim(\mathsf{W})$

Corollary 2.19.1. Let V be a vector space over F. Then V is isomorphic to F^n if and only if $\dim(\mathsf{V}) = n$.

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 $\Phi : \mathcal{L}(\mathsf{V}, \mathsf{W}) \to \mathsf{M}_{m\times n}(F)$, defined by $\Phi(\mathsf{T}) = [\mathsf{T}]_\beta^\gamma$ for $\mathsf{T} \in \mathcal{L}(\mathsf{V}, \mathsf{W})$, is an isomorphism.

Corollary 2.20.1. Let V and W be finite-dimensional vector spaces of dimensions n and m, respectively. The, $\mathcal{L}(\mathsf{V}, \mathsf{W})$ is finite dimensional of dimension mn.

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

Theorem 2.21. For any finite-dimensional vector space V with ordered basis β, ϕ_β is 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 $Q = [\mathsf{I}_V]_{\beta'}^\beta$. Then

(a) Q is invertible.

(b) For any $V \in \mathsf{V}$, $[v]_\beta = Q[v]_{\beta'}$

Note: The matrix $Q = [\mathsf{I}_V]_{\beta'}^{\beta}$ defined in Theorem 2.22 is called **change of coordinate matrix**. Because of part (b) of the theorem, we say that Q **changes β'−coordinates into β-coordinates**.

Linear operator: a linear transformations that map a vector space V into itself is called a **linear operator** on V

Theorem 2.23. Let T be a linear operator on 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

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

Corollary 2.23.1. Let $A \in \mathsf{M}_{n\times n}(F)$, and let γ be an ordered basis for F^n. Then $[\mathsf{L}_A]_\gamma = Q^{-1}AQ$, 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 $\mathsf{M}_{n\times n}(F)$. We say that V is **similar** to A if there exists an invertible matrix Q such that $B = Q^{-1}AQ$.

2.6 Dual Spaces *

2.7 Homogeneous Linear Differential Equations with Constant Coefficients *

3 Elementary Matrix Operations and Systems of Linear Equations

3.1 Elementary Matrix Operations and Elementary Matrices

Definition. 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].

Note: 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, type 2,** or **type 3** according to whether the elementary operation performed on I_n is a type 1,2, or 3 operation respectively.

Theorem 3.1. Let $A \in \mathsf{M}_{m\times n}(F)$, 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 I_n [I_n].

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 $A \in \mathsf{M}_{m\times n}(F)$, we define the **rank** of A, denoted rank(A), to be the rank of the linear transformation $\mathsf{L}_A : \mathsf{F}^n \to \mathsf{F}^m$

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

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

(a) $\text{rank}(AQ) = \text{rank}(A)$.

(b) $\text{rank}(PA) = \text{rank}(A)$.

and therefore,

(c) $\text{rank}(PAQ) = \text{rank}(A)$.

Corollary 3.4.1. Elementary row and column operations on a matrix are rank-preserving.

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 $m \times n$ 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 3.6.1. Let A be an $m \times n$ matrix of rank r. Then there exist invertible matrices B and C of sizes $m \times m$ and $n \times n$, respectively, such that $D = BAC$ where

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

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

Corollary 3.6.2. Let A be an $n \times n$ matrix. Then

(a) $\text{rank}(A^t) = \text{rank}(A)$.

(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 matrix.

Corollary 3.6.3. Every invertible matrix is a product of elementary matrices.

Theorem 3.7. Let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ and $\mathsf{U} : \mathsf{W} \to \mathsf{Z}$ be linear transformations on finiite-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).

Definition. Let A and B be $m \times n$ and $n \times p$ matrices, respectively. By the **augmented matrix** $(A|B)$, we mean the $m \times (n+p)$ 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.

3.3 Systems of Linear Equations – Theoretical Aspects

The system of equations (S)

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{1n}x_n &= b_n \end{aligned}$$

where a_{ij} and b_i ($1 \leq i \leq m$ and $1 \leq j \leq n$) are scalers 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

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

is called the **coeffiecient matrix** of the system (S)

Let

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

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

$$Ax = b.$$

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

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

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

Definition. 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 $\mathsf{K} = \mathsf{N}(\mathsf{L}_A)$; hence K is a subspace of F^n of dimension $n - \text{rank}(\mathsf{L}_A) = n - \text{rank}(A)$

Corollary 3.8.1. $m \leq n$, the system $Ax = 0$ has a nonzero solution.

Note: $Ax = 0$ is referred to 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 $Ax = b$

$$K = \{s\} + \mathsf{K}_\mathsf{H} = \{s + k : k \in \mathsf{K}_\mathsf{H}\}$$

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.

Note: 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$(A|b)$

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 3.12. 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 $CAx = Cb$ is equivalent to $Ax = b$

Corollary 3.12.1. 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.

Theorem 3.13. Gaussian elimination transforms any matrix into its reduced row echelon form.

Note: an arbitrary solution of the form

$$s = s_0 + t_1u_1 + t_2u_2 + \ldots + t_{n-r}u_{n-r},$$

where r is the number of nonzero rows in A' $r \leq m$. The preceding equation is called a **general solution of the system** $Ax = b$,here u_i's and are the basis for the solution set and the t_i's are the coeffiecient and s_0 is a particular solution to the original system.

Theorem 3.14. 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 genreal solution obtained by the procedure above is of the form

$$s = s_0 + t_1u_1 + t_2u_2 + \ldots + t_{n-r}u_{n-r},$$

then $\{u_1, u_2, \ldots, u_{n-4}\}$ is a basis for the solution set of the corresponding homogeneous system, and s_0 is a solution to the original system.

Theorem 3.15. Let A be an $m \times n$ matrix of rank r, where $r \geq 0$, 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 $i = 1, 2, \ldots, r$, 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 $k = 1, 2, \ldots, n$, if column k of B is $d_1e_1 + d_2e_2 + \ldots + d_re_r$, then column k of A is $d_1a_{j_1} + d_2a_{j_2} + \ldots + d_na_{j_n}$

Corollary 3.15.1. The reduced row echelon form of a matrix is unique.

4 Determinants

4.1 Determinants of Order 2

Theorem 4.1.

Theorem 4.2.

4.2 Determinants of Order n

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

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

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

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

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

is called **cofactor** of the entry of A in row i, column j entry of A, we can express the formula for the determinant of A as

$$\det(A) = A_{11}c_{11} + A_{12}c_{12} + \ldots + A_{1n} + c_{1n}$$

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

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

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

Corollary 4.3.1. If $A \in \mathsf{M}_{n\times n}(F)$ has a row consisting entirely of zeros, then $\det(A) = 0$

Lemma. Let $B \in \mathsf{M}_{n\times n}(F)$, 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 $A \in \mathsf{M}_{n\times n}(F)$, then for any integer i $(1 \leq i \leq n)$,

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

Corollary 4.4.1. If $A \in \mathsf{M}_{n\times n}(F)$ has two identical rows, then $\det(A) = 0$

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

Theorem 4.6. Let $A \in \mathsf{M}_{n\times n}(F)$, 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 4.6.1. If $A \in \mathsf{M}_{n\times n}(F)$ has rank less than n, then $\det(A) = 0$

4.3 Properties of Determinants

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

Corollary 4.7.1. A matrix $A \in \mathsf{M}_{m\times m}(F)$ 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 $A \in \mathsf{M}_{n\times n}(F)$, $\det(A^t) = det(A)$

Theorem 4.9 (Cramer's Rule)**.** Let $Ax = b$ be a 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 $(k = 0, 1, \ldots, n)$,

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

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

Note: it is possible to interpret the determinant of a matrix $A \in \mathsf{M}_{n\times n}(R)$ geometrically. 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 a 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.

4.4 Summary – Important Facts about Determinants

4.5 A Characterization of the Determinant

5 Diagonalization

5.1 Eigenvalues and Eigenvectors

Definition. 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 $[\mathsf{T}]_\beta$ is a diagonal matrix. A square matrix A is called **diagonalizable** if L_A is diagonalizable.

Definition. Let T be a linear operator on a vector space V. A nonzero vector $v \in \mathsf{V}$ is called an **eigenvector** of T if there exists a scalar λ such that $\mathsf{T}(v) = \lambda v$. The scalar λ is called the **eigenvalue** corresponding to the eigenvector v.
Let A be in $\mathsf{M}_{n\times n}(F)$. A nonzero vector $v \in \mathsf{F}^n$ 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.

Note: the words characteristic polynomial and proper vector are also used in place of eignvector.

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 $D = [\mathsf{T}]_\beta$, then D is a diagonal matrix and D_{jj} is the eigenvalue corresponding to v_j for $1 \leq j \leq n$.

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

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

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

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

Theorem 5.3. Let $A \in \mathsf{M}_{n\times n}(F)$.

(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 $v \in \mathsf{V}$ is an eigenvector of T corresponding to λ if and only if $v \neq 0$ and $v \in \mathsf{N}(\mathsf{T} - \lambda\mathsf{I})$

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 5.5.1. 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 $\mathsf{P}(F)$ **splits over** F if there are scalars $c, a_1, a_2, \ldots, a_n$ (not necessarily distinct) in F such that

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

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

Definition. Let λ be an eigenvalue of a linear operator or a 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 $\mathsf{E}_\lambda = \{x \in \mathsf{V} : \mathsf{T}(x) = \lambda x\} = \mathsf{N}(\mathsf{T} - \lambda\mathsf{I}_\mathsf{V})$. 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 $1 \leq \dim(\mathsf{E}_\lambda) \leq m$.

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 $v_i \in \mathsf{E}_{\lambda_i}$, the eigenspace corresponding to λ_i. If

$$v_1 + v_2 + \ldots + v_k = 0,$$

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 $S = S_1 \cup S_2 \cup \ldots \cup S_k$ 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 distinct eigenvalues of T. Then

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

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

Direct Sums

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

$$\{v_1 + v_2 + \ldots v_k : v_i \in \mathsf{W}_i \text{ for } 1 \leq i \leq k\}$$

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

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

and

$$\mathsf{W}_j \cap \sum_{i \neq j} \mathsf{W_i} = \{0\} \quad \text{for each } j \ (1 \leq j \leq k).$$

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

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

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

(c) Each vector $v \in \mathsf{V}$ can be uniquely written as $v = v_1 + v_2 + \cdots + v_k$, where $v_i \in \mathsf{W}_i$

(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 $i = 1, 2, \ldots, k$, 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 *

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 $\mathsf{T}(\mathsf{W}) \subseteq \mathsf{W}$, that is, if $\mathsf{T}(v) \in \mathsf{W}$ for all $v \in \mathsf{W}$.

The subspace $\mathsf{W} = \operatorname{span}(\{x, \mathsf{T}(x), \mathsf{T}^2(x), \ldots\})$ is called **T-cyclic subspace** of V **generated by** x.

Theorem 5.12. 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 T_W divides the characteristic polynomial of T.

Theorem 5.13. 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 $v \in \mathsf{V}$. Let $k = \dim(\mathsf{W})$. Then

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

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

Theorem 5.14 (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 $f(\mathsf{T}) = \mathsf{T}_0$, the zero transformation. That is, T "satisfies" its characteristic equation.

Corollary 5.14.1 (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 5.15. Let T be a linear operator on a finite-dimensional vector space V, and suppose that $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2 \oplus \cdots \oplus \mathsf{W}_k$, where W_i is a T-invariant subspace of V for each i $(1 \le i \le k)$. Suppose that $f_i(t)$ is the characteristic polynomial of $\mathsf{T}_{\mathsf{W}_i}$ $(1 \le i \le k)$. Then $f_1(t) \cdot f_2(t) \cdot \cdots \cdot f_k(t)$ is the characteristic polynomial of T.

Definition. Let $B_1 \in \mathsf{M}_{m\times m}(F)$, and let $B_2 \in \mathsf{M}_{n\times n}(F)$. 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

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

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 5.16. Let T be a linear operator on a finite-dimensional vector space V, and let $\mathsf{W}_1, \mathsf{W}_2, \ldots, \mathsf{W}_k$ be T-invariant subspaces of V such that $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2 \oplus \cdots \oplus \mathsf{W}_k$. 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 $A = [\mathsf{T}]_\beta$ and $B_i = [\mathsf{T}_{\mathsf{W}_i}]_{\beta_i}$ for $i = 1, 2, \ldots, k$. Then $A = B_1 \oplus B_2 \oplus \cdots \oplus B_k$.

6 Inner Product Spaces

6.1 Inner Product and Norms

Definition. Let V be a vector space over a field F. And **inner product** on V is a function that assigns, to every ordered pair of vectors x and y in V, a scaler 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 $x \neq 0$.

Note: For $x = (a_1, a_2, \ldots, a_n)$ and $y = (b_1, b_2, \ldots, b_n)$ in F^n, define

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

the **standard inner product**

Definition. Let $A \in \mathsf{M}_{m\times n}(F)$. 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.

Let $\mathsf{V} = \mathsf{M}_{n\times n}(F)$, and define $\langle A, B \rangle = \operatorname{tr}(B^*A)$ for $A, B \in \mathsf{V}$ 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 **complex inner product space,** whereas if $F = R$, we call V **real inner product space.**

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

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

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

(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 $x \in \mathsf{V}$, then $y = z$.

Note 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 $x \in \mathsf{V}$, 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 $x, y \in \mathsf{V}$ and $c \in F$, the following statements are true.

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

(b) $||x|| = 0$ if and only if $x = 0$. In ansy 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: The process of multiplying a nonzero vector by the reciprocal of its length is called **normalizing.**

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 $S = \{v_1, v_2, \ldots, v_k\}$ be an orthogonal subset of V consisting of nonzero vectors. If $y \in \text{span}(S)$, then

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

Corollary 6.3.1. If, in addition to hypotheses of Theorem 6.3, S is orthonormal and $y \in \text{span}(S)$, then

$$y = \sum_{i_1}^{k} \langle y, v_i\rangle v_i.$$

Corollary 6.3.2. Let V be an inner product space, and let S be an orthogonal subset of V consisting of nonzero vectors. Then S is a linearly independent.

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

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{||v_j||^2} v_j \quad \text{for } 2 \leq k \leq n. \tag{1}$$

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

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

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

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

Corollary 6.5.1. Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, v_2, \ldots, v_n\}$. Let T be a linear operator with an orthonormal basis $\beta = \{v_1, v_2, \ldots, v_n\}$. Let T be a linear operator on V, and let $A = [\mathsf{T}]_\beta$. Then for any i and j, $A_{ij} = \langle \mathsf{T}(v_j), v_i \rangle$

Definition. Let β be an orthonormal subset (possibly infinite) of an inner product space V, and let $x \in \mathsf{V}$. 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, $S^\perp = \{x \in \mathsf{V} : \langle x, y \rangle \ = 0 \text{ for all } y \in S\}$. 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 $y \in \mathsf{V}$. Then there exist unique vectors $u \in \mathsf{W}$ and $z \in \mathsf{W}^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, v_2, \ldots v_k\}$ is orthonormal basis for W, then

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

Corollary 6.6.1. 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 $x \in \mathsf{W}$, $||y - x|| \geq ||y - u||$, and this inequality is an equality if an only if $x = u$

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

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

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

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

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

6.3 The Adjoint of a Linear Operator

For a linear operator T on an inner product space V, we now define a related linear operator on V called adjoint of T whose matrix representation with respect to any orthonormal basis β for V is $[\mathsf{T}]_\beta^*$

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

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 $\mathsf{T}^* : \mathsf{V} \to \mathsf{V}$ such that $\langle \mathsf{T}(x), y \rangle$ for all $x, y \in \mathsf{V}$. Furthermore, T^* is linear.

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

$$[\mathsf{T}^*]_\beta = [\mathsf{T}]_\beta^*.$$

Corollary 6.10.1. Let A be an $n \times n$ matrix. Then $\mathsf{L}_{A^*} = (\mathsf{L}_A)^*$

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

(a) $(\mathsf{T} + \mathsf{U})^* = \mathsf{T}^* + \mathsf{U}^*$;

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

(c) $(\mathsf{TU})^* = \mathsf{U}^*\mathsf{T}^*$;

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

(e) $\mathsf{I}^* = \mathsf{I}$.

Corollary 6.11.1. 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) $I^* = I$.

Lemma. Let $A \in \mathsf{M}_{m\times n}(F), x \in \mathsf{F}^n$, and $y \in \mathsf{F}^m$. Then

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

Lemma. Let $A \in \mathsf{M}_{m\times n}(F)$. Then $\operatorname{rank}(A^*A) = \operatorname{rank}(A)$.

Corollary 6.11.2. If A is an $m \times n$ matrix such that $\operatorname{rank}(A) = n$, then A^*A is invertible.

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

Minimal solution: a solution s to $Ax = b$ is called a **minimal solution** if $||s|| \leq ||u||$ for all other solutions u.

Theorem 6.13. Let $A \in \mathsf{M}_{m\times n}(F)$ and $b \in \mathsf{F}^m$. Suppose that $Ax = b$ is consistent. Then the following statements are true.

(a) There exists exactly one minimal solution s of $Ax = b$, and $s \in \mathsf{R}(\mathsf{L}_{A^*})$.

(b) The vector s is the only solution to $Ax = b$ that lies in $\mathsf{R}(\mathsf{L}_{A^*})$; that is, if u satisfies $(AA^*)u = b$,then $s = A^*u$

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 $[\mathsf{T}]_\beta$ is upper triangular.

Definition. Let V be an inner product space, and let T be a linear operator on V. We say that T is **normal** if $\mathsf{T}\mathsf{T}^* = \mathsf{T}^*\mathsf{T}$. 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) $||\mathsf{T}(x)|| = ||T^*(x)||$ for all $x \in \mathsf{V}$.

(b) $\mathsf{T} - c\mathsf{I}$ 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 $\mathsf{T}(x) = \lambda x$, then $\mathsf{T}^*(x) = \overline{\lambda}(x.)$

(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.

Definition. Let T be a linear operator on an inner product space V. We say that T is **self-adjoint (Hermitian)** if $\mathsf{T} = \mathsf{T}^*$. 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 self adjoint if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.

6.5 Unitary and Orthogonal Operators and Their Matrices

Definition. Let T be a linear operator on a finite-dimensional inner product space V (over F). If $||\mathsf{T}(\mathsf{x})|| = ||x||$ for all $x \in \mathsf{V}$, we call T is a **unitary operator** if $F = C$ and an **orthogonal operator** if $F = R$.
Note: 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) $\mathsf{TT}^* = \mathsf{T}^*\mathsf{T} = \mathsf{I}$.

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

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

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

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

Lemma. Let U be a self-adjoint operator on a finite-dimensional inner product space V. If $\langle x, \mathsf{U}(x)\rangle = 0$ for all $x \in \mathsf{V}$, then $\mathsf{U} = \mathsf{T}_0$

Corollary 6.18.1. 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 6.18.2. 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 $\mathsf{T}(x) = x$ for all and $\mathsf{T}(x) = -x$ for all $x \in \mathsf{L}^\perp$

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

For a complex normal [real symmetric] matrix A, there exists an orthonormal basis β for F^n consisting of eigenvectors of A. Hence is similar to diagonal matrix D. By the corollary to Theorem 2.23, the matrix Q whose columns are the vectors in β is such tat $D = Q^{-1}AQ$. 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.

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 (Shur)**.** Let $A \in \mathsf{M}_{n\times n}(F)$ 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 *

Definition. Let V be a real inner product space. A function $f : \mathsf{V} \to \mathsf{V}$ is called a **rigid motion** if

$$||f(x) - f(y)|| = ||x - y||$$

A class of rigid motions are translations. A function $g : \mathsf{V} \to \mathsf{V}$, where V is a real inner product space, is called a **translation if there exists a vector** $v_0 \in \mathsf{V}$ such taht $g(x) = x + v_0$ for all $x \in \mathsf{V}$. We say that g is a translation by v_0.

Theorem 6.22. Let $f : \mathsf{V} \to \mathsf{V}$ be a rigid motion on a finite-dimensional real inner product space V. Then there exists a unique orthogonal operator T on V and a unique translation g on V such that $f = g \circ \mathsf{T}$

Orthogonal Operators on R^2

Theorem 6.23. Let T be an orthogonal operator on R^2, and let $A = [\mathsf{T}]_\beta$, where β is the standard ordered basis for R^2. Then exactly one of the following conditions is satisfied:

(a) T is a rotation, and $\det(A) = 1$.

(b) T is a reflection about a line through the origin, and $\det(A) = -1$

Corollary 6.23.1. Any rigid motion on R^2 is either a rotation followed by a translation or a reflection about a line through the origin followed by a translation.

6.6 Orthogonal Projections and the Spectral Theorem

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

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 $\mathsf{T}^2 = \mathsf{T} - \mathsf{T}^*$

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 $\mathsf{W_i}$. Then the following statements are true.

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

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

(c) $\mathsf{T}_i\mathsf{T}_j = \delta_{ij}\mathsf{T_i}$ for $1 \leq i, j \leq k$.

(d) $\mathsf{I} = \mathsf{T_1} + \mathsf{T_2} + \cdots + \mathsf{T}_k$.

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

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

Corollary 6.25.1. If $F = C$, then T is normal if and only if $\mathsf{T}^* = g(\mathsf{T})$ for some polynomial g.

Corollary 6.25.2. If $F = C$, then T is unitary if and only if T is normal and $|\lambda| = 1$ for every eigenvalue λ of T.

Corollary 6.25.3. If $F = C$ and T is normal, then T is self-adjoint if and only if every eigenvalue of T is real.

Corollary 6.25.4. Let T be as in the spectral theorem with spectral decomposition $\mathsf{T} = \lambda_1\mathsf{T}_1 + \lambda_2\mathsf{T}_2 + \cdots + \lambda_k T_k$. Then each T_j is a polynomial in T.

6.7 The Singular Value Decomposition and the Pseudoinverse *

6.8 Bilinear and Quadratic Forms *

6.9 Einstein's Special Theory of Relativity *

6.10 Conditioning and Rayleigh Quotient *

6.11 The Geometry of Orthogonal Operators *

7 Canonical Forms

7.1 The Jordan Canonical Form I

The definition of eigenspace is expanded to generalized eigenspace. The ordered bases whose union is an ordered basis β for V such that

$$[\mathsf{T}]_\beta \begin{pmatrix} A_1 & O & \dots & O \\ O & A_2 & \dots & O \\ \vdots & \vdots & & \vdots \\ O & O & \dots & A_k \end{pmatrix},$$

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

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

for some eigenvalue λ of T. Such a matrix A_i is called a **Jordan block** corresponding to λ, and the matrix $[\mathsf{T}]_\beta$ is called a **Jordan canonical form** for T. We also say that the ordered basis β is a **Jordan canonical basis** for T.

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 $(\mathsf{T} - \lambda\mathsf{I})^p(x) = 0$ 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

$$\mathsf{K}_\lambda = \{x \in \mathsf{V} : (\mathsf{T} - \lambda\mathsf{I})^p(x) = 0 \text{ for some positive interger } p\}$$

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 $\mathsf{T} - \mu\mathsf{I}$ 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) $\dim(\mathsf{K}_\lambda) \leq m$.

(b) $\mathsf{K}_\lambda = \mathsf{N}((\mathsf{T} - \lambda\mathsf{I})^{\mathsf{m}})$

Theorem 7.3. Let T be a linear operator on a finite-dimensional vector space V such taht the characteristic polynomial of T splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of T. Then, for every $x \in \mathsf{V}$, there exists vectors $v_i \in \mathsf{K}_{\lambda_i}, 1 \leq i \leq k$, 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 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) $\dim(\mathsf{K}_{\lambda_i}) = m_i$ for all i.

Corollary 7.4.1. 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 $\mathsf{E}_\lambda = \mathsf{K}_\lambda$ for every eigenvalue λ of T.

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

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

is called a **cycle of generalized eigenvectors** of T corresponding to λ. The vectors $(\mathsf{T} - \lambda\mathsf{I})^{p-1}(x)$ 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, $\mathsf{W} = \operatorname{span}(\gamma)$ is T-invariant, and $[\mathsf{T_W}]_\gamma$ 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 $\gamma = \displaystyle\bigcup_{i=1}^{q} \gamma_i$ is linearly independent.

Corollary 7.6.1. 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 7.7.1. 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 $A \in \mathsf{M}_{n\times n}(F)$ be such that characteristics 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 7.7.2. 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 visualize each of the matrices A_i and ordered basis 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}. Suppose that β_i is a disjoint union of cycles of generalized eigenvectors $\gamma_1, \gamma_2, \ldots, \gamma_{n_i}$ with lengths $p_1 \leq p_2 \leq \cdots \leq p_{n_i}$ respectively. The dot of $\mathsf{T_i}$ contains one dot for each vector in β_i, and the dots are configured according to the following rules.

1. The array consists of n_i columns (one column for each cycle).
2. Counting from left to right, the jth column consists of the p_j dots that correspond to the vectors of γ_j starting with the initial vector at top and continuing down to the end vector.

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 $\mathsf{N}((\mathsf{T} - \lambda_i\mathsf{I})^{\mathsf{r}})$. Hence the number of dots in the first r rows of the dot diagram equals $\operatorname{nullity}(\mathsf{T}-\lambda_i\mathsf{I})^r)$.

Corollary 7.9.1. 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) $r_1 = \dim(\mathsf{V}) - \operatorname{rank}(\mathsf{T} - \lambda_i \mathsf{I})$,

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

Corollary 7.10.1. 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.

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.

7.3 The minimal Polynomial

Definition. Let T be a linear operator on a finite-dimensional vector space. A polynomial $p(t)$ is called a **minimal polynomial** of T if $p(t)$ is a monic polynomial of least positive degree for which $p(\mathsf{T}) = \mathsf{T}_0$.

Theorem 7.12. Let $p(t)$ be a minimal polynomial of a linear operator T on a finite-dimensional vector space V.

(a) For any polynomial $g(\mathsf{T}) = \mathsf{T}_0$, then $p(t)$ divides $g(t)$. In particular, $p(t)$ divides the characteristic polynomial of T.

(b) The minimal polynomial of T is unique.

Definition. Let $A \in \mathsf{M}_{n\times n}(F)$. The **minimal polynomial** $p(t)$ of A is the monic polynomial of least positive degree for which $p(A) = O$

Theorem 7.13. Let T be a linear operator on a finite-dimensional vector space V, and let β be an ordered basis for V. Then the minimal polynomial of T is the same as the minimal polynomial of $[\mathsf{T}]_\beta$.

Corollary 7.13.1. For any $A \in \mathsf{M}_{n\times n}(F)$, the minimal polynomial of A is the same as the minimal polynomial of L_A.

Theorem 7.14. Let T be a linear operator on a finite-dimensional vector space V, and let $p(t)$ be the minimal polynomial of T. A scalar λ is an eigenvalue of T if and only if $p(\lambda) = 0$ Hence the characteristic polynomial and the minimal polynomial of T have the smae zeros.

Corollary 7.14.1. Let T be a linear operator on a finite-dimensional vector space V with minimal polynomial $p(t)$ and characteristic polynomial $f(t)$. Suppose that $f(t)$ factors as

$$f(t) = (\lambda_1 - t)^{n_1}(\lambda_2 - t)^{n_2} \cdots (\lambda_k - t)^{n_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct eigenvalues of T. Then there exist integers $m_1, m_2, \ldots, m_k$ such that $1 \leq m_i \leq n_i$ for all i and

$$p(t) = (t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2} \cdots (t - \lambda_k)^{m_k}.$$

Theorem 7.15. Let T be a linear operator on an n-dimensional vector space V such that V is a T-cycle subspace of itself. Then the characteristic polynomial $f(t)$ and the minimal polynomial $p(t)$ have the same degree, and hence $f(t) = (-1)^n p(t)$.

Theorem 7.16. Let T be a linear operator on a finite-dimensional vector space V. Then T is diagonalizable if and only if the minimal polynomial of T is of the form

$$p(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_k)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct eigenvalues of T.

7.4 The Rational Canonical Form *