LINEAR ALGEBRA
ROBERT GRONE AND JOHN DUSEL

CONTENTS

1. Vector Spaces 3
2. Basis and Dimension 5
3. Linear Transformations 7
4. Matrices of Linear Transformations 11
5. Eigenvalues and Eigenvectors 14
6. The Characteristic Polynomial 17
7. The Cayley-Hamilton Theorem 21
8. Similarity 25
9. Diagonalization 27
10. Complex Numbers 31
11. The Adjoint of a Complex Matrix 34
12. Jordan Form 36
13. Inner Product Spaces and Euclidean Norm 39
14. Orthogonal and Orthonormal Bases, Orthogonal Projections 44
15. Orthogonal Subspaces and Orthogonal Complements 46
16. Some Special Complex Matrices 48
17. Unitary Similarity and Schur’s Theorem 51
18. Unitary Diagonalization and the Spectral Theorem 55
19. Unitary Equivalence and the Singular Value Decomposition 59
Index 61

Date: Fall 2005.
These are the typed notes from San Diego State University’s Math 524 course given Fall 2005. All content is due to Dr. Robert Grone, any typographical errors are due to JMD.
1. Vector Spaces

Let $F$ denote an arbitrary field.

Example.

(1) $F = \mathbb{R}$, the real numbers,
(2) $F = \mathbb{C}$, the complex numbers,
(3) $F = \mathbb{Q}$, the rational numbers,
(4) $F = \mathbb{Z}_2 = \{0, 1\}$, with $1 + 1 = 0$, the binary field.

We defer the axioms and definitions for fields to Math 521AB.

Definition. A set $V(\neq \emptyset)$ is a vector space over $F$ if $V$ has two operations, vector addition and scalar multiplication, that satisfy the following properties. Let $u, v, w$ denote arbitrary elements of $V$ and $\alpha, \beta$ denote arbitrary elements of $F$. The vector space axioms (which we shall not dwell on) are as follows:

(1) $u + v \in V$
(2) $u + v = v + u$
(3) $(u + v) + w = u + (v + w)$
(4) there exists $0 \in V$ such $u + 0 = u$
(5) there exists $-u \in V$ such that $u + (-u) = 0$
(6) $\alpha u \in V$
(7) $\alpha(u + v) = \alpha u + \alpha v$
(8) $(\alpha + \beta)u = \alpha u + \beta u$
(9) $\alpha(\beta u) = (\alpha \beta)u$
(10) $1u = u$

Example.

(1) $V = M_{m,n}(F) = \text{all } m\text{-by-}n \text{ matrices with entries in } F$
(2) $V = F^n = M_{n,1} = \text{all } n\text{-tuples with entries in } F$
(3) $V = F[x] = \text{all polynomials in } x \text{ with coefficients in } F$
(4) $V = C_0(\mathbb{R}) = \text{all continuous functions from } \mathbb{R} \text{ to } \mathbb{R}$

Let $V$ be a vector space over $F$. A subset $\emptyset \neq W \subseteq V$ which is also a vector space over $F$ is called a subspace of $V$.

Theorem 1.1 (Subspace Test). Let $V$ be a vector space over $F$ and let $\emptyset \leq W \subseteq V$. Then $W$ is a subspace of $V$ iff $W$ satisfies:

(1) $u + v \in W$ for all $u, v \in W$ (i.e., $W$ is closed under addition)
(2) $\alpha u \in W$, all $\alpha \in F$, all $u \in W$ (i.e., $W$ is closed under scalar multiplication)
Remark. In some texts, the above two conditions are used to define the notion of subspace. In any event, this theorem provides a concrete way of showing that a subset of $V$ is (or isn’t) a subspace of $V$.

Example.
(1) Let $V = M_n(\mathbb{R}) = \text{all } n \times n \text{ real matrices}$ and let
$$W = \{ A \in V : A^T = A \}$$
be the set of all symmetric matrices in $V$. Then $W$ is a subspace of $V$.
(2) Let $V = \mathbb{R}[x]$ and $W = \{ p(x) \in \mathbb{R}[x] : p(x) = 0 \}$. Then $W$ is a subspace of $V$.
(3) The **nullspace** of $A \in M_{m,n}(\mathbb{R})$ is
$$\{ v \in \mathbb{R}^n : Av = 0 \}.$$ For any $A \in M_{m,n}(\mathbb{R})$, the nullspace of $A$ is a subspace of $\mathbb{R}^n$.

Let $v_1, \ldots, v_n \in V$, where $V$ is a vector space. Any vector that can be written as
$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$for some $\alpha_1, \ldots, \alpha_n \in F$ is referred to as a **linear combination** of $v_1, \ldots, v_n$. The set of all linear combinations of $v_1, \ldots, v_n$ is the **span** of $v_1, \ldots, v_n$.

**Theorem 1.2.** Let $V$ be a vector space and let $v_1, \ldots, v_n \in V$. Then span$\{v_1, \ldots, v_n\}$ is a subspace of $V$.

Remark.
(1) If $v_1, \ldots, v_n \in W$ and $W$ is a subspace of $V$, then span$\{v_1, \ldots, v_n\} \subseteq W$.
(2) $v_i \in \text{span}\{v_1, \ldots, v_n\}$, all $i = 1, \ldots, n$.

Suppose $v_1, \ldots, v_n \in V$ and that $W$ is a subspace of $V$. We say that
$$v_1, \ldots, v_n \, \text{span} \, W$$or
$$\{v_1, \ldots, v_n\} \, \text{spans} \, W$$iff
$$W = \text{span}\{v_1, \ldots, v_n\}.$$In this case we say that $\{v_1, \ldots, v_n\}$ is a **spanning set** for $W$. 

2. Basis and Dimension

Let \( V \) be a vector space and let \( v_1, \ldots, v_n \in V \). We say that \( v_1, \ldots, v_n \) are linearly independent if the only solution to
\[
\alpha_1 v_1 + \cdots + \alpha_n v_n = 0
\]
is
\[
\alpha_1 = \cdots = \alpha_n = 0.
\]
(This is the trivial solution). Otherwise, we say that \( v_1, \ldots, v_n \) are linearly dependent.

Example. The vectors
\[
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}, \quad \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix}, \quad \begin{bmatrix}
2 \\
3 \\
4
\end{bmatrix}
\]
are linearly dependent since
\[
1 \begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix} + 1 \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} - 1 \begin{bmatrix}
2 \\
3 \\
4
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\]

Definition. We say that a set \( S = \{v_1, \ldots, v_n\} \) is a basis of \( V \) if and only if

(1) \( S \) spans \( V \)
(2) \( S \) is linearly independent.

Example.

(1) The standard basis of \( \mathbb{R}^2 \) is \( \{e_1, e_2\} \) where
\[
e_1 = \begin{bmatrix}
1 \\
0
\end{bmatrix}, \quad e_2 = \begin{bmatrix}
0 \\
1
\end{bmatrix}.
\]

(2) Two nonzero vectors in \( \mathbb{R}^2 \), say \( u \) and \( v \), form a basis of \( \mathbb{R}^2 \) if and only if \( v \) is not a multiple of \( u \).

(3) The standard basis of \( \mathbb{R}^3 \) is \( \{e_1, e_2, e_3\} \) where
\[
e_1 = \begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}, \quad e_2 = \begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}, \quad e_3 = \begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}.
\]

(4) The standard basis of \( M_2(\mathbb{R}) \) is
\[
\{E_{11}, E_{12}, E_{13}, E_{14}\} = \left\{ \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}, \begin{bmatrix}
0 & 1 \\
0 & 0
\end{bmatrix}, \begin{bmatrix}
0 & 0 \\
1 & 0
\end{bmatrix}, \begin{bmatrix}
0 & 0 \\
0 & 1
\end{bmatrix} \right\}.
\]
We will assume the following axiom: Every vector space has a basis.

The next theorem gives an alternate characterization of bases (plural of basis)

**Theorem 2.1.** Let $V$ be a vector space and suppose $S = \{v_1, \ldots, v_n\} \subseteq V$. Then $S$ is a basis of $V$ if and only if every $v \in V$ can be written uniquely as a linear combination of $v_1, \ldots, v_n$.

The next theorem is fairly easy to prove in $\mathbb{R}^n$ using Gaussian Elimination.

**Theorem 2.2.** Let $S = \{v_1, \ldots, v_n\}$ be a basis of a vector space $V$. Then any set of $m > n$ vectors in $V$ is linearly dependent.

**Corollary 2.3.** Let $V$ be a vector space with basis $S = \{v_1, \ldots, v_n\}$. Then any basis of $V$ has exactly $n$ vectors.

**Definition.** The dimension of a vector space $V$ is the number of vectors in a basis of $V$.

**Example.**
1. $\dim F^n = n$
2. $\dim M_{m,n}(F) = mn$
3. If $A \in M_{m,n}(\mathbb{R})$, then the nullity of $A$ equals the dimension of the nullspace of $A$.
4. If $A \in M_{m,n}(\mathbb{R})$, then the span of the rows of $A$ is the rowspace of $A$ and the rank of $A$ is the dimension of the rowspace of $A$.

The next result has been called the “fundamental theorem of linear algebra” by some. This is certainly an exaggeration, but it is inarguably an important result.

**Theorem 2.4 (Sylvester’s Law of Nullity).** Suppose $A \in M_{m,n}(F)$. Then $\text{rank}(A) + \text{nullity}(A) = n$.

**Remark.** A complete understanding of Gaussian Elimination is sufficient to prove the above result. Roughly speaking;

\[
\begin{align*}
n &= \text{number of variables}, \\
\text{rank}(A) &= \text{number of lead variables} \\
\text{nullity}(A) &= \text{number of free variables}.
\end{align*}
\]
3. Linear Transformations

Let $U$ and $V$ be vector spaces over a field $F$. We say that a mapping

$$T : U \to V$$

is a linear transformation if $T$ satisfies

1. $T(u_1 + u_2) = T(u_1) + T(u_2)$; all $u_1, u_2 \in U$
2. $T(\alpha u) = \alpha T(u)$; all $u \in U$, all $\alpha \in F$.

We use $\mathcal{L}(U, V)$ to denote the set of all linear transformations from $U$ to $V$.

**Theorem 3.1.** Let $T_1, T_2 \in \mathcal{L}(U, V)$. Then

1. $T_1 + T_2 \in \mathcal{L}(U, V)$
2. $\alpha T_1 \in \mathcal{L}(U, V)$ for all $\alpha \in F$.

**Proof.** Let $u_1, u_2 \in U$ and $\alpha \in F$. Then

$$(T_1 + T_2)(u_1 + u_2) = T_1(u_1 + u_2) + T_2(u_1 + u_2)$$
$$= T_1(u_1) + T_1(u_2) + T_2(u_1) + T_2(u_2)$$
$$= T_1(u_1) + T_2(u_1) + T_1(u_2) + T_2(u_2)$$
$$= (T_1 + T_2)(u_1) + (T_1 + T_2)(u_2)$$

and

$$(T_1 + T_2)(\alpha u) = T_1(\alpha u) + T_2(\alpha u)$$
$$= \alpha T_1(u) + \alpha T_2(u)$$
$$= \alpha [T_1(u) + T_2(u)]$$
$$= \alpha (T_1 + T_2)(u).$$

\[ \square \]

**Theorem 3.2.** Let $U, V, W$ be vector spaces over $F$. Suppose $T_1 \in \mathcal{L}(U, V)$ and $T_2 \in \mathcal{L}(V, W)$, so that

$$T_2 \circ T_1 : U \to W.$$

Then $T_2 \circ T_1 \in \mathcal{L}(U, W)$.

**Proof.** Let $u_1, u_2 \in U$ and $\alpha \in F$. Then

$$T_2 \circ T_1(u_1 + u_2) = T_2(T_1(u_1 + u_2))$$
$$= T_2(T_1(u_1) + T_1(u_2))$$
$$= T_2(T_1(u_1)) + T_2(T_1(u_2))$$
$$= T_2 \circ T_1(u_1) + T_2 \circ T_1(u_2).$$
and

\[
T_2 \circ T_1(\alpha u) = T_2(T_1(\alpha u)) \\
= T_2(\alpha T_1(u)) \\
= \alpha T_2(T_1(u)) \\
= \alpha T_2 \circ T_1(u).
\]

\[\square\]

**Theorem 3.3.** Let \( T \in \mathcal{L}(U, V) \) be invertible with inverse

\[ T^{-1} : V \to U. \]

Then \( T^{-1} \in \mathcal{L}(V, U) \).

**Proof.** Homework. \[\square\]

Suppose \( T \in \mathcal{L}(U, V) \). We define the **nullspace** or **kernel** of \( T \) as

\[ \ker(T) = \{ u \in U : T(u) = 0 \}. \]

We define the **range** of \( T \) as

\[ \text{range}(T) = \{ T(u) : u \in U \} \subseteq V. \]

**Theorem 3.4.** Let \( T \in \mathcal{L}(U, V) \). Then

1. \( \ker(T) \) is a subspace of \( U \)
2. \( \text{range}(T) \) is a subspace of \( V \).

**Proof of Theorem 4, part (2).** Suppose \( v_1, v_2 \in \text{range}(T) \) and \( \alpha \in F \) (to show \( v_1 + v_2 \in \text{range}(T) \) and \( \alpha v_1 \in \text{range}(T) \)). Say

\[ v_1 = T(u_1) \quad \text{and} \quad v_1 = T(u_2) \]

where \( u_1, u_2 \in U \). Then

\[ v_1 + v_2 = T(u_1) + T(u_2) = T(u_1 + u_2), \]

so \( v_1 + v_2 \in \text{range}(T) \) and

\[ \alpha v_1 = \alpha T(u_1) = T(\alpha u_1), \]

so \( \alpha v_1 \in \text{range}(T) \). \[\square\]

Suppose \( T \in \mathcal{L}(U, V) \). In view of Theorem 4, we may make the following definitions

\[ \text{nullity}(T) = \dim(\ker(T)) \quad \text{and} \quad \text{rank}(T) = \dim(\text{range}(T)). \]

Our next theorem is a more general version of a theorem from Math 254.
Theorem 3.5. Let \( T : \mathcal{L}(U, V) \), and suppose \( \dim(U) = n < \infty \). Then
\[
\text{rank}(T) + \text{nullity}(T) = n
\]
or
\[
\text{rank}(T) + \text{nullity}(T) = \dim(U).
\]

Before proving Theorem 5, we present two useful lemmas which will simplify the proof.

Lemma 3.6. Let \( T \in \mathcal{L}(U, V) \). Suppose \( u_1, \ldots, u_k \in U \) and \( \alpha_1, \ldots, \alpha_k \in F \). Then
\[
T(\alpha_1 u_1 + \cdots + \alpha_k u_k) = \alpha_1 T(u_1) + \cdots + \alpha_k T(u_k).
\]

Lemma 3.7. Let \( T \in \mathcal{L}(U, V) \). Suppose \( u_1, \ldots, u_k \in U \) and that \( T(u_1), \ldots, T(u_k) \) are linearly independent in \( V \). Then \( u_1, \ldots, u_k \) are linearly independent in \( U \).

Proof of Lemma 6. Use induction on \( k \).

Proof of Lemma 7. Suppose \( \alpha_1 u_1 + \cdots + \alpha_k u_k = 0 \). (To show \( \alpha_1 = \cdots = \alpha_k = 0 \).) Then
\[
0 = T(0) = T(\alpha_1 u_1 + \cdots + \alpha_k u_k) = \alpha_1 T(u_1) + \cdots + \alpha_k T(u_k).
\]

Then, since \( T(u_1), \ldots, T(u_k) \) are linearly independent in \( V \), \( \alpha_1 = \cdots = \alpha_k = 0 \). Hence \( u_1, \ldots, u_k \) are linearly independent in \( U \).

Proof of Theorem 5. Let \( k = \text{nullity}(T) = \dim(\ker(T)) \), and let \( n = \dim(U) \). Since \( \ker(T) \) is a subspace of \( U \), \( k \leq n \). Let \( \{u_1, \ldots, u_k\} \) be a basis of \( \ker(T) \). Extend this to a basis of \( V \), say \( \{u_1, \ldots, u_k, u_{k+1}, \ldots, u_n\} \).

Any \( u \in U \) can be written uniquely as
\[
 u = \alpha_1 u_1 + \cdots + \alpha_k u_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_n u_n.
\]

For such a \( u \),
\[
 T(u) = \alpha_1 T(u_1) + \cdots + \alpha_k T(u_k) + \alpha_{k+1} T(u_{k+1}) + \cdots + \alpha_n T(u_n)
 = \alpha_1 0 + \cdots + \alpha_k 0 + \alpha_{k+1} T(u_{k+1}) + \cdots + \alpha_n T(u_n)
 = \alpha_{k+1} T(u_{k+1}) + \cdots + \alpha_n T(u_n)
\]

This shows that \( \{T(u_{k+1}), \ldots, T(u_n)\} \) spans \( \text{range}(T) \). We claim that this set is also linearly independent. (This will finish the proof, since then \( \{T(u_{k+1}), \ldots, T(u_n)\} \) is a basis of \( \text{range}(T) \), so \( \text{rank}(T) = n - k = \dim(U) - \text{nullity}(T) \).)

Suppose, for a contradiction, that
\[
 \alpha_{k+1} T(u_{k+1}) + \cdots + \alpha_n T(u_n) = 0.
\]
Then
\[ T(\alpha_{k+1}u_{k+1}) + \cdots + T(\alpha_n u_n) = 0, \]
so
\[ \alpha_{k+1}u_{k+1} + \cdots + \alpha_n u_n \in \ker(T). \]
Since \( \{u_1, \ldots, u_k\} \) is a basis of \( \ker(T) \), \( \alpha_{k+1}u_{k+1} + \cdots + \alpha_n u_n \) can be written as a linear combination of \( \{u_1, \ldots, u_k\} \). Unless \( \alpha_{k+1} = \cdots = \alpha_n = 0 \), this would contradict our assumption that \( \{u_1, \ldots, u_k, u_{k+1}, \ldots, u_n\} \) is linearly independent. Hence \( \alpha_{k+1} = \cdots = \alpha_n = 0 \), which completes the proof. \( \square \)

Remark. Theorem 5 is valid even when \( \dim(U) = \infty \).
4. Matrices of Linear Transformations

Suppose \( U \) is a vector space over a field \( F \) with \( \dim(U) = n < \infty \) and that \( B = \{u_1, \ldots, u_n\} \) is a basis of \( U \). Recall that any \( u \in U \) can be written uniquely as

\[
u = \alpha_1 u_1 + \cdots + \alpha_n u_n,
\]

which allows us to define the **coordinate vector of \( u \) with respect to \( B \)** as

\[
(u)_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in F^n.
\]

**Theorem 4.1.** Let \( U \) and \( B \) be as described above; and let \( u, v \in U \), \( \alpha \in F \). Then

1. \((u + v)_B = (u)_B + (v)_B\)
2. \((\alpha u)_B = \alpha (u)_B\).

*Proof of Theorem 1, part (1).* Let

\[
(u)_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in F^n
\]

and

\[
(v)_B = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \in F^n
\]

so

\[
u = \alpha_1 u_1 + \cdots + \alpha_n u_n, \quad v = \beta_1 u_1 + \cdots + \beta_n u_n.
\]

Then

\[
u + v = (\alpha_1 + \beta_1) u_1 + \cdots + (\alpha_n + \beta_n) u_n
\]

which implies

\[
(u + v)_B = \begin{bmatrix} \alpha_1 + \beta_1 \\ \vdots \\ \alpha_n + \beta_n \end{bmatrix}
\]

\[
= \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}
\]

\[
= (u)_B + (v)_B.
\]

Proof of part (2) is left to the reader.
Corollary 4.2. The map from $U$ to $R^n$ defined by $u \mapsto (u)_B$ is a linear transformation.

Remark. The map described in Corollary 2 is one-to-one and onto, and hence is invertible.

Our goal is to describe a matrix which corresponds to $T \in L(U, V)$ when $\dim(U) = n < \infty$ and $\dim(V) = m < \infty$. Suppose $B_1 = \{u_1, \ldots, u_n\}$ and $B_2 = \{v_1, \ldots, v_m\}$ are bases of $U$ and $V$, respectively. We define the matrix representation of $T$ relative to $B_1$ and $B_2$ by

$$[T]_{B_2}^{B_1} = \begin{bmatrix} (T(u_1))_{B_2} & \cdots & (T(u_n))_{B_2} \end{bmatrix}$$

so that $[T]_{B_2}^{B_1} \in M_{m,n}(F)$.

With this definition, it is easy to prove the following result.

Theorem 4.3. Let $U$ and $V$ have bases $B_1$ and $B_2$ respectively, with $|B_1| = n < \infty$ and $|B_2| = m < \infty$. If $T_1, T_3 \in L(U, V)$ and $\alpha \in F$, then

1. $[T_1 + T_2]_{B_1}^{B_2} = [T_1]_{B_1}^{B_2} + [T_2]_{B_1}^{B_2}$
2. $[\alpha T_1]_{B_1}^{B_2} = \alpha [T_1]_{B_1}^{B_2}$

Proof of Theorem 3, part (2). By definition,

$$\alpha [T_1]_{B_1}^{B_2} = \alpha \begin{bmatrix} (T_1(u_1))_{B_2} & \cdots & (T_1(u_n))_{B_2} \end{bmatrix} = \begin{bmatrix} \alpha (T_1(u_1))_{B_2} & \cdots & \alpha (T_1(u_n))_{B_2} \end{bmatrix} = [\alpha T_1]_{B_1}^{B_2}. \quad \square$$

The next result gives the connection between coordinate vectors and matrix representations.

Theorem 4.4. Let $U$ and $V$ be vector spaces over $F$ with bases $B_1$ and $B_2$. Assume $|B_1| = n = \dim(U)$ and $|B_2| = m = \dim(V)$. Let $T \in L(U, V)$ and $u \in U$. Then

$$[T]_{B_1}^{B_2}(u)_{B_1} = (T(u_1))_{B_2}.$$  

Proof. Let $u = \alpha_1 u_1 + \cdots + \alpha_n u_n$, so that

$$(u)_{B_1} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$
Then

\[
[T]_{B_1}^{B_2}(u)_{B_1} = \left( (T(u_1))_{B_2} \cdots (T(u_n))_{B_2} \right) \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}
\]

\[= \alpha_1(T(u_1))_{B_2} + \cdots + \alpha_n(T(u_n))_{B_2} = (\alpha_1T(u_1))_{B_2} + \cdots + (\alpha_nT(u_n))_{B_2} = (T(\alpha_1u_1 + \cdots + \alpha_nu_n))_{B_2} = (T(u_1))_{B_2}. \]

\[\square\]

Theorem 4 says something like this. If we choose bases \(B_1\) and \(B_2\) for \(U\) and \(V\), then

\(U\) turns into \(F^n\)

\(V\) turns into \(F^m\), and

\(\mathcal{L}(U,V)\) turns into \(M_{m,n}(F)\).

**Theorem 4.5.** Let \(U, V, W\) be finite dimensional vector spaces over \(F\) of dimension \(n, m,\) and \(p\) respectively, and suppose \(B_1, B_2, B_3\) are bases of \(U, V, W\), respectively. Let \(T_1 \in \mathcal{L}(U,V)\) and \(T_2 \in \mathcal{L}(V,W)\) so that \(T_2 \circ T_1 \in \mathcal{L}(U,W)\). Then

\[\left[ T_2 \right]^{B_2}_{B_1} \left[ T_1 \right]^{B_2}_{B_1} = \left[ T_2 \circ T_1 \right]^{B_2}_{B_1}.\]

**Remark.** Note that \(\left[ T_1 \right]^{B_2}_{B_1}\) is \(m \times n\), \(\left[ T_2 \right]^{B_2}_{B_1}\) is \(p \times n\), so that the matrix multiplication in the statement of Theorem 5 is at least plausible.

**Proof of Theorem 5.** Assume \(B_1 = \{u_1, \ldots, u_n\}\). By definition,

\[\left[ T_2 \circ T_1 \right]^{B_2}_{B_1} = \left[ \left( T_2 \circ T_1(u_1) \right)_{B_3} \cdots \left( T_2 \circ T_1(u_n) \right)_{B_3} \right] = \left[ \left( T_2(T_1(u_1)) \right)_{B_3} \cdots \left( T_2(T_1(u_n)) \right)_{B_3} \right] = \left[ \left( T_2 \right)_{B_2}^{B_3} \left( T_1(u_1) \right)_{B_2} \cdots \left( T_2 \right)_{B_2}^{B_3} \left( T_1(u_n) \right)_{B_2} \right] = \left[ T_2 \right]_{B_2}^{B_3} \left[ T_1 \right]_{B_2}^{B_3} = \left[ T_2 \circ T_1 \right]^{B_2}_{B_1},\]

where the equalities come from the definition of \(\left[ T_1 \right]^{B_2}_{B_1}\), the definition of \(T_2 \circ T_1\), Theorem 4, matrix multiplication, and the definition of \(\left[ T_1 \right]^{B_2}_{B_1}\) respectively. \(\square\)
5. Eigenvalues and Eigenvectors

Let $M_n(F)$ denote the set of all $n$-by-$n$ matrices with entries in the field $F$. We say that $\lambda \in F$ is an eigenvalue of $A \in M_n(F)$ if there exists $0 \neq v \in F^n$ such that

$$Av = \lambda v.$$ 

**Example.** Let

$$A = \begin{bmatrix} 2 & 3 \\ 6 & -1 \end{bmatrix} \in M_2(\mathbb{R}).$$

Then $-4$ is an eigenvalue of $A$ since

$$A \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -4 \\ 8 \end{bmatrix} = -4 \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$ 

If $\lambda$ is an eigenvalue of $A$ and $Av = \lambda v$ ($v \neq 0$), we refer to $v$ as an eigenvector of $A$ corresponding to the eigenvalue $\lambda$. Note that $v = 0$ always satisfies $Av = \lambda v$ for any $\lambda$, but $v = 0$ is excluded from being an eigenvector by definition.

Suppose $\lambda$ is an eigenvalue of $A$. The set of eigenvectors corresponding to $\lambda$ and $A$, together with $v = 0$, is called the eigenspace corresponding to $\lambda$ and $A$. In other words, the eigenspace corresponding to $\lambda$ and $A$ is

$$\{v \in F^n : Av = \lambda v\}.$$ 

**Theorem 5.1.** Let $\lambda \in F$ be an eigenvalue of $A \in M_n(F)$. Then the corresponding eigenspace is a (nonzero) subspace of $F^n$.

**Proof.** By definition,

$$\{v \in F^n : Av = \lambda v\} = \{v \in F^n : Av - \lambda v = 0\}$$

$$= \{v \in F^n : Av - \lambda I_n v = 0\}$$

$$= \{v \in F^n : (A - \lambda I_n)v = 0\}$$

$$= \text{nullspace}(A - \lambda I_n).$$

Since the nullspace of any $n$-by-$n$ matrix is a subspace of $F^n$, the eigenspace is a subspace of $F^n$. The eigenspace is a nonzero subspace since we are assuming that there exists a (nonzero) eigenvector corresponding to $\lambda$ and $A$. 

**Remark.** Looking at the proof of Theorem 1 tells us how to find a basis for an eigenspace. Specifically, if $\lambda$ is an eigenvalue of $A$, we can compute RREF$(A - \lambda I_n)$.
Example. Let

\[ A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}. \]

Given that \( \lambda = 2 \) is an eigenvalue of \( A \), find a basis for the corresponding eigenspace.

Here, \( A - \lambda I_n = A - 2I_3 \). So

\[
\text{RREF}(A - 2I_3) = \text{RREF} \left( \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right)
= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\]

Hence a general eigenvector corresponding to \( \lambda = 2 \) and \( A \) is

\[ v = \alpha \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \]

A basis for the corresponding eigenspace is

\[ \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}. \]

Suppose \( \lambda \) is an eigenvalue of \( A \). The dimension of the corresponding eigenspace is called the geometric multiplicity of \( \lambda \) as an eigenvalue of \( A \). In the previous example,

\[ \text{geom. mult.}(2) = 2. \]

In more generality, if \( \lambda \) is an eigenvalue of \( A \), then

\[
\text{geom. mult.}(\lambda) = \dim \{ v \in F^n : Av = \lambda v \}
= \dim(\text{nullspace}(A - \lambda I_n))
= \text{nullity}(A - \lambda I_n).
\]

Suppose we know that a particular \( \lambda \) is an eigenvalue of \( A \in M_n(F) \). The previous remarks and examples show how to find a basis for the corresponding eigenspace by computing \( \text{RREF}(A - \lambda I_n) \). We ask the following question.

Question. Given \( A \in M_n(F) \), how do we decide which scalars \( \lambda \in F \) are eigenvalues of \( A \)?
Theorem 5.2. Let \( A \in M_n(F) \) and \( \lambda \in F \). Then
\[ \lambda \text{ is an eigenvalue of } A \]
if and only if
\[ \det(A - \lambda I_n) = 0. \]

Proof. By definition,
\[ \lambda \text{ is an eigenvalue of } A \]
if and only if
\[ Av = \lambda v \text{ has a solution } v \neq 0 \]
if and only if
\[ Av - \lambda v = 0 \text{ has a solution } v \neq 0 \]
if and only if
\[ Av - \lambda I_n v = 0 \text{ has a solution } v \neq 0 \]
if and only if
\[ (A - \lambda I_n)v = 0 \text{ has a solution } v \neq 0 \]
if and only if
\[ \det(A - \lambda I_n) = 0. \]

We refer to the equation
\[ \det(A - \lambda I_n) = 0 \]
as the characteristic equation of \( A \).

Example. Use Theorem 2 to find all eigenvalues of
\[ A = \begin{bmatrix} 5 & 7 \\ 1 & 11 \end{bmatrix}. \]

Solution
\[ 0 = \det(A - \lambda I_2) \]
\[ = \det \left( \begin{bmatrix} 5 - \lambda & 7 \\ 1 & 11 - \lambda \end{bmatrix} \right) \]
\[ = (5 - \lambda)(11 - \lambda) - 7 \cdot 1 \]
\[ = \lambda^2 - 16\lambda + 48 \]
\[ = (\lambda - 12)(\lambda - 4). \]

Hence \( \lambda_1 = 12 \) and \( \lambda_2 = 4 \) are the only eigenvalues of \( A \). (We may use the method in Theorem 1 to find the corresponding eigenvectors to \( \lambda_1 \) and \( \lambda_2 \), if desired.)
6. THE CHARACTERISTIC POLYNOMIAL

Let $A \in M_n(F)$. We have seen that the eigenvalues of $A$ are the solutions to the characteristic equation of $A$,
\[ \det(A - \lambda I_n) = 0. \]
(Or, in some books, $\det(\lambda I_n - A) = 0$.) If we think of $\lambda$ as an unknown (or indeterminate), then the characteristic polynomial of $A$ is
\[ \det(A - \lambda I_n). \]

**Example.** For
\[ A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}, \]
the characteristic equation of $A$ is
\[ \lambda^2 - 4\lambda - 5 = 0 \]
and the characteristic polynomial of $A$ is
\[ \lambda^2 - 4\lambda - 5 \]
(details left to the reader).

**Example.** If
\[ A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \]
then the characteristic polynomial of $A$ is
\[
\det(A - \lambda I_2) = \det\left( \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} \right) \\
= (a - \lambda)(d - \lambda) - bc \\
= \lambda^2 - (a + d)\lambda + (ad - bc) \\
= \lambda^2 - \text{trace}(A)\lambda + \det(A).
\]

This example shows that the characteristic polynomial of any 2-by-2 matrix is a quadratic polynomial in the unknown $\lambda$. Hence any 2-by-2 matrix can have at most 2 distinct eigenvalues. For a general $n \geq 2$, we have the following:

**Theorem 6.1.** Let $A \in M_n(F)$. Then $\det(A - \lambda I_n)$ has the form
\[ (-1)^n\lambda^n + (-1)^{n-1}c_1\lambda^{n-1} + \cdots - c_{n-1}\lambda + c_n. \]

The proof of this result requires some knowledge of determinants. As a consequence we have the following:

**Corollary 6.2.** If $A \in M_n(F)$, then $A$ can have no more than $n$ distinct eigenvalues.
Proof of corollary 2. The eigenvalues of $A$ are roots of the characteristic polynomial, which by Theorem 1 is a polynomial of degree $n$. Since a polynomial of degree $n$ can have at most $n$ distinct roots, $A$ has at most distinct eigenvalues. □

The next result will express the coefficients in the characteristic polynomial of $A$ in terms of the entries of $A$. First we need some notation. For any $1 \leq k \leq n$, $A$ has $k$-by-$k$ submatrices which are obtained by specifying $k$ rows and $k$ columns of $A$. The following example illustrates the idea and introduces some notation.

Example. Let

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}.$$  

The following are examples of 2-by-2 submatrices of $A$:

1. If we use rows 1 and 2 and columns 3 and 4, we obtain

$$A[1, 2|3, 4] = \begin{bmatrix} 3 & 4 \\ 7 & 8 \end{bmatrix}.$$  

2. If we use rows 2 and 3 and columns 2 and 3, we obtain

$$A[2, 3|2, 3] = \begin{bmatrix} 6 & 7 \\ 10 & 11 \end{bmatrix}.$$  

3. If we use rows 1 and 4 and columns 2 and 4, we obtain

$$A[1, 4|2, 4] = \begin{bmatrix} 2 & 4 \\ 14 & 16 \end{bmatrix}.$$  

4. If we use rows 1 and 3 and columns 1 and 3, we obtain

$$A[1, 3|1, 3] = \begin{bmatrix} 1 & 3 \\ 9 & 11 \end{bmatrix}.$$  

To denote a general $k$-by-$k$ submatrix of an $n$-by-$n$ matrix $A$, we use the notation

$$A[\alpha|\beta]$$

where $\alpha$ represents a sequence of $k$ rows and $\beta$ represents a sequence of $k$ columns. By a $k$-by-$k$ subdeterminant of an $n$-by-$n$ matrix $A$ is meant any

$$\det A[\alpha|\beta]$$

where $A[\alpha|\beta]$ is a $k$-by-$k$ submatrix of $A$. If $\alpha = \beta$, we refer to $A[\alpha|\alpha]$ as a principal submatrix.
and
\[ \det A[\alpha|\alpha] \] as a principal subdeterminant.

**Example.** Let
\[
A = \begin{bmatrix}
1 & 2 & 3 & 4 \\
5 & 6 & 7 & 8 \\
9 & 10 & 11 & 12 \\
13 & 14 & 15 & 16
\end{bmatrix}.
\]
Then \( A \) has \( \binom{4}{2} = 6 \) principal 2-by-2 submatrices. These are:
\[
\begin{align*}
A[1,2|1,2] &= \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix} \\
A[1,3|1,3] &= \begin{bmatrix} 1 & 3 \\ 9 & 11 \end{bmatrix} \\
A[1,4|1,4] &= \begin{bmatrix} 1 & 4 \\ 13 & 16 \end{bmatrix} \\
A[2,3|2,3] &= \begin{bmatrix} 6 & 7 \\ 10 & 11 \end{bmatrix} \\
A[2,4|2,4] &= \begin{bmatrix} 6 & 8 \\ 14 & 16 \end{bmatrix} \\
A[3,4|3,4] &= \begin{bmatrix} 11 & 12 \\ 15 & 16 \end{bmatrix}.
\end{align*}
\]

**Theorem 6.3.** Suppose \( A \in M_n(F) \) has characteristic polynomial
\[
\det(A - \lambda I_n) = (-1)^n \lambda^n + (-1)^{n-1} c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n.
\]
Then
\[
c_k = \sum_{|\alpha|=k} \det A[\alpha|\alpha].
\]
i.e., the sum of all the \( k \)-by-\( k \) principal subdeterminants of \( A \) equals \( c_k \).

**Remark.**
(1) If \( k = 1 \), then \( c_1 = a_{11} + a_{22} + \cdots + a_{nn} = \text{trace}(A) \).
(2) If \( k = n \), then \( c_n = \det(A) \).

Suppose \( A \in M_n(F) \), and that the characteristic polynomial of \( A \) factors into \( n \) linear terms (not necessarily distinct). Say
\[
\det(A - \lambda I_n) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n),
\]
where \( \lambda_1, \ldots, \lambda_n \in F \). Using just polynomial multiplication properties, we can express \( c_1, \ldots, c_n \) as follows.
Theorem 6.4. Let $A$ and $c_1, \ldots, c_n$ be as in Theorem 3 and above. Then

1. $c_1 = \lambda_1 + \cdots + \lambda_n = \text{trace}(A)$,
2. $c_n = \lambda_1 \ldots \lambda_n = \text{det}(A)$,
3. $c_k = \text{the sum of all products of } k \text{ of } \lambda_1, \ldots, \lambda_n$.

(This last is $E_k(\lambda_1, \ldots, \lambda_n)$, the $k$th elementary symmetric function of $\lambda_1, \ldots, \lambda_n$.)

Suppose $\lambda_0$ is an eigenvalue of $A \in M_n(F)$. Recall that we have defined the **geometric multiplicity of $\lambda_0$ as an eigenvalues of $A$** as the dimension of the corresponding eigenspace,

$$\text{geom. mult.}(\lambda_0) = \dim\{v \in F^n : Av = \lambda_0 v\}.$$ 

We define the **algebraic multiplicity of $\lambda_0$ as an eigenvalue of $A$** as the multiplicity of $\lambda_0$ as a root of the characteristic polynomial of $A$.

**Example.** Let

$$A = \begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix}.$$ 

Then

$$\text{geom. mult.}(\lambda_0) = 1$$

and

$$\text{alg. mult.}(\lambda_0) = 2.$$ 

Theorem 6.5. Suppose $\lambda_0$ is an eigenvalue of $A \in M_n(F)$. Then

$$1 \leq \text{geom. mult.}(\lambda_0) \leq \text{alg. mult.}(\lambda_0) \leq n.$$
7. The Cayley-Hamilton Theorem

For any polynomial
\[ p(x) = a_0 + a_1 x + \cdots + a_m x^m \in F[x] \]
and any matrix \( A \in M_n(F) \), we may define
\[ p(A) = a_0 + a_1 A + \cdots + a_m A^m \in M_n(F). \]

Example.
(1) If \( p(x) = x^k \), then \( p(A) = A^k \).
(2) If \( p(x) = 2 + x^2 \), then \( p(A) = 2 + A^2 \).
(3) If \( p(x) = 2 - 3x + 4x^3 \), then \( P(A) = 2 - 3A + 4A^3 \).
(4) If \( p(x) = 2 - 3x + 4x^3 \) and
\[
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},
\]
then
\[
p(A) = \begin{bmatrix} 27 & 34 \\ 51 & 78 \end{bmatrix}.
\]
(The reader should check this.)

Recall: If \( p(x), q(x) \in F[x] \) and \( \alpha \in F \), we define

(1) \((pq)(x) = p(x)q(x)\)
(2) \((p + q)(x) = p(x) + q(x)\)
(3) \((\alpha p)(x) = \alpha p(x)\)

**Theorem 7.1.** Suppose \( p(x), q(x) \in F[x], \alpha \in F, \) and \( A \in M_n(F) \).
Then
\[
(1) \ (pq)(A) = p(A)q(A)
(2) \ (p + q)(A) = p(A) + q(A)
(3) \ (\alpha p)(A) = \alpha p(A).
\]

**Lemma 7.2.** Suppose \( S, A \in M_n(F) \) and \( S \) invertible. Then

(1) \( S^{-1}A^kS = (S^{-1}AS)^k \), all \( k = 1, 2, 3, \ldots \)
(2) \( S^{-1}(\alpha A)S = \alpha S^{-1}AS \), all \( \alpha \in F \)
(3) \( (A^k)^T = (A^T)^k \), all \( k = 1, 2, 3, \ldots \).

**Theorem 7.3.** Suppose \( S, A \in M_n(F) \) with \( S \) invertible. Then

(1) \( p(S^{-1}AS) = S^{-1}p(a)S \); all \( p(x) \in F[x] \)
(2) \( p(A^T) = p(A)^T \); all \( p(x) \in F[x] \).
Proof of Theorem 3. For part (i), if \( p(x) = a_0 + a_1 x + \cdots + a_m x^m \), then
\[
S^{-1} p(A) S = S^{-1} (a_0 I_n + a_1 A + \cdots + a_m A^m) S
\]
\[
= S^{-1} a_0 I_n S + S^{-1} a_1 AS + \cdots + S^{-1} a_m A^m S
\]
\[
= a_0 S^{-1} I_n S + a_1 S^{-1} AS + \cdots + a_m S^{-1} A^m S
\]
\[
= a_0 I_n + a_1 S^{-1} AS + \cdots + a_m S^{-1} A^m S
\]
\[
= p(S^{-1} AS).
\]
As for part (ii),
\[
p(A)^T = (a_0 I_n + a_1 A + \cdots + a_m A^m)^T
\]
\[
= (a_0 (I_n)^T + a_1 A^T + \cdots + a_m (A^m)^T)
\]
\[
= (a_0 (I_n)^T + a_1 A^T + \cdots + a_m (A^T)^m)
\]
\[
= p(A^T).
\]
\( \square \)

We say that \( p(x) \in F[x] \) annihilates \( A \in M_n(F) \) if and only if \( p(A) = 0 \).

Example. Let
\[
A = \begin{bmatrix}
1 & 3 \\
2 & 1
\end{bmatrix}
\]
and \( p(x) = x^2 - 2x - 5 \) \((= \det(A - xI_2))\). Then
\[
p(A) = A^2 - 2A - 5I_2
\]
\[
= \begin{bmatrix}
7 & 6 \\
4 & 7
\end{bmatrix} - 2 \begin{bmatrix}
1 & 3 \\
2 & 1
\end{bmatrix} - 5 \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]
\[
= \begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}.
\]

Example. \( p(x) = x^2 - (a + d)x + (ad - bc) \) annihilates
\[
A = \begin{bmatrix}
a & b \\
c & d
\end{bmatrix}.
\]

Theorem 7.4. Suppose \( S, A \in M_n(F) \) with \( S \) invertible, and that \( p(x) \in F[x] \). Then
\( (1) \) \( p(A) = 0 \) if and only if \( p(S^{-1} AS) = 0 \)
\( (2) \) \( P(A) = 0 \) if and only if \( p(A^T) = 0 \).

Theorem 7.5. Suppose \( p(x), q(x), s(x) \in F[x] \) and that \( p(x) \) and \( q(x) \) annihilate \( A \in M_n(F) \). Then
\( (1) \) \( p(x) + q(x) \) annihilates \( A \)
\( (2) \) \( p(x)s(x) \) annihilates \( A \).
Proof of Theorem 5.

1. \( p(A) + q(A) = 0 + 0 = 0. \)
2. \( p(A)s(A) = 0 \cdot s(A) = 0. \)

\[ \square \]

Theorem 7.6. Let \( A \in M_n(F) \). Then there exists \( 0 \neq p(x) \in F[x] \) which annihilates \( A \).

Proof. Since \( \dim M_n(F) = n^2 \), \( I_n, A, \ldots, A^{n^2} \) are linearly dependent. Hence there exists non-trivial scalars \( c_0, c_1, \ldots, c_{n^2} \) such that

\[
 c_0I_n + c_1A + \cdots + c_{n^2}A^{n^2} = 0.
\]

Then

\[
 p(x) = c_0 + c_1x + \cdots + c_{n^2}x^{n^2}
\]

is a nonzero polynomial that annihilates \( A \).

\( \square \)

Let \( A \in M_n(F) \). A **minimal** (or **minimum**) polynomial for \( A \) is a monic polynomial of least degree which annihilates \( A \).

Theorem 7.7. Let \( A \in M_n(F) \). Then

1. \( A \) has a unique minimum polynomial, say \( m_A(x) \)
2. For any \( p(x) \in F[x] \),

\[
 p(A) = 0 \quad \text{if and only if} \quad m_A(x) \mid p(x).
\]

Proof.

1. By Theorem 6, there exists nontrivial polynomials which annihilate \( A \). Hence there exist monic polynomials of least degree which annihilate \( A \). Suppose \( m_1(x) \) and \( m_2(x) \) are monic polynomials of least degree which annihilate \( A \). If \( m_1(x) \neq m_2(x) \), then \( m_1(x) - m_2(x) \) is a nonzero polynomial of lesser degree. Furthermore, by Theorem 5, \( m_1(x) - m_2(x) \) annihilates \( A \). This contradicts our assumption that \( m_1(x) \) and \( m_2(x) \) had minimal degree among the polynomials which annihilate \( A \).

2. \((\Leftarrow)\) Assume \( m_A(x) \mid p(x) \). Say \( m_a(x)q(x) = p(x) \). Then

\[
 p(A) = m_a(A)q(A) = 0q(A) = 0.
\]

\((\Rightarrow)\) Suppose \( p(A) = 0 \). Assume, for contradiction, that \( m_A(x) \) does not divide \( p(x) \). Then write

\[
 p_A(x) = m_A(x)q(x) + r(x)
\]

where \( r(x) \neq 0 \) and \( \deg r(x) < \deg m_A(x) \). Then

\[
 r(x) = p_A(x) - m_A(x)q(x)
\]
and

\[ r(A) = p(A) - m_A(A)q(A) \]
\[ = 0 - 0q(A) = 0. \]

This contradicts the fact that \( \deg m_A(x) \) is minimal among all polynomials which annihilate \( A \).

\[ \square \]

\textbf{Theorem 7.8} (Cayley-Hamilton). \textit{Let} \( A \in M_n(F) \) \textit{have minimal polynomial} \( m_A(x) \) \textit{and characteristic polynomial} \( ch_A(x) \). \textit{Then} \( m_A(x) \mid ch_A(x) \).

\textit{Remark.} An equivalent statement is \( ch_A(x) = 0 \) or “the characteristic polynomial of \( A \) is an annihilating polynomial for \( A \)”. 
8. Similarity

We say that $A, B \in M_n(F)$ are similar if and only if

$$S^{-1}AS = B$$

for some invertible $S \in M_n(F)$.

**Theorem 8.1** (Similarity is an Equivalence Relation). Suppose $A, B, C \in M_n(F)$. Then

1. $A$ is similar to $A$
2. $A$ is similar to $B$ if and only if $B$ is similar to $A$.
3. If $A$ is similar to $B$ and $B$ is similar to $C$, then $A$ is similar to $C$.

*Remark.* On justification for studying matrices comes from the fact that we can say that two matrices are similar if and only if they represent the same linear transformation on an $n$-dimensional vector space with respect to two different bases. In this case, the matrix $S$ is called a “transition matrix” or “change of basis matrix”.

The notion of similarity has applications in matrix theory to studying the eigenvalues and eigenvectors of a matrix, and in defining $f(A)$, where $f(x)$ is a function (not necessarily a polynomial). The functions $f(A) = e^{\alpha A}$ are of particular importance in differential equations.

**Theorem 8.2.** Suppose $A, B \in M_n(F)$ are similar. Then

1. $A^k$ is similar to $B^k$, all $k = 1, 2, \ldots$.
2. $A^T$ is similar to $B^T$.
3. If $A$ and $B$ are invertible, then $A^{-1}$ is similar to $B^{-1}$.
4. For any polynomial $p(x)$, $p(A)$ is similar to $p(B)$.

**Theorem 8.3.** Suppose $A, B \in M_n(F)$ are similar. Then $A$ and $B$ have the same

1. rank
2. determinant
3. characteristic polynomial
4. eigenvalues
5. algebraic multiplicities of eigenvalues
6. geometric multiplicities of eigenvalues
7. minimal polynomial

Following are examples of how to prove some of the parts of the previous three theorems. You will be asked to prove some parts yourself in homework #3.
Proof of Theorem 6.1 part (2). Assume $B$ is similar to $A$. Say $S^{-1}AS = B$. Solving for $A$ yields
\[ A = SBS^{-1} = (S^{-1})^{-1}B(S^{-1}) \]
which shows $A$ is similar to $B$. □

Proof of Theorem 6.2 part (1). Assume $B$ is similar to $A$. Say $S^{-1}AS = B$. Then
\[ S^{-1}A^kS = (S^{-1}AS)^k, \quad \text{all } k = 1, 2, \ldots \]
shows that $A^k$ and $B^k$ are similar for all $k = 1, 2, \ldots$. □

Proof of Theorem 6.2 part (3). Assume $B$ is similar to $A$. Say $S^{-1}AS = B$. Then
\[
ch_B(x) = \det(B - xI_n) \\
= \det(S^{-1}AS - xI_n) \\
= \det(S^{-1}(A - xI_n)S) \\
= \det(S^{-1})\det(A - xI_n)\det(S) \\
= \frac{1}{\det(S)}\det(A - xI_n)\det(S) \\
- \det(A - xI_n) \\
= ch_A(x).
\]
Proof. Assume $A$ and $B$ are similar. Then, by Theorem 6.3 part (3), $A$ and $B$ have the same characteristic polynomials. Hence $A$ and $B$ have the same eigenvalues. □

Proof. Assume $A$ and $B$ are similar. Then, by Theorem 6.3 part (3), $A$ and $B$ have the same characteristic polynomial. Since the algebraic multiplicities of eigenvalues are defined in terms of the corresponding characteristic polynomial, they are the same for $A$ and $B$. □
9. Diagonalization

Suppose $A \in M_n(F)$. We say that $A$ is **diagonalizable** (over $F$) if and only if there exists invertible $S$ and diagonal $D$ (in $M_n(F)$) such that

$$S^{-1}AS = D.$$ 

In other words, $A$ is similar to a diagonal matrix $D \in M_n(F)$.

**Theorem 9.1.** Suppose $S, A, D \in M_n(F)$ with $S$ invertible, $D$ diagonal, and

$$S^{-1}AS = D.$$ 

Then, the diagonal entries of $D$ are the eigenvalues of $A$ (with multiplicities).

**Proof.** Since $A$ and $D$ are similar, they have the same eigenvalues, and the eigenvalues of $D$ are its diagonal entries. \qed

**Remark.** Theorem 1 answers part of the question of finding an $S$ and $D$ for a given $A \in M_n(F)$ that will satisfy $S^{-1}AS = D$. Namely, finding $D$ amounts to finding the eigenvalues of $A$.

**Example.**

(1) If

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

then $S^{-1}AS = D$ for

$$S = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

(2) If

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

then $S^{-1}AS = D$ for

$$S = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

(3) If

$$A = \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix},$$

then $S^{-1}AS = D$ for

$$S = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}.$$
Example. Consider
\[ A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \]
which has eigenvalues \( \pm i \). There cannot be a real matrix \( S \) such that
\[ S^{-1}AS = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \]
so \( A \) is not diagonalizable over \( \mathbb{R} \). However
\[ A = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} \]
does satisfy the above equation, thus \( A \) is diagonalizable over \( \mathbb{C} \) but not \( \mathbb{R} \).

Example. For
\[ A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \]
there is no matrix of any kind where \( S^{-1}AS \) is diagonal. This matrix is not diagonalizable over \( F \) for any \( F \).

Theorem 1 told how to find a candidate for \( D \) in the equation
\[ S^{-1}AS = D \] for a fixed \( A \). The next result describes how to find \( S \). A closer look shows that it also gives a method to decide whether \( A \) is diagonalizable.

**Theorem 9.2.** Let \( S, A, D \in M_n(F) \) with \( S \) invertible and \( D = \text{diag}(\lambda_1, \ldots, \lambda_n) \). Let \( v_1, \ldots, v_d \) be the columns of \( S \),
\[ S = [v_1 \ v_2 \ \ldots \ v_n]. \]
Then
\[ S^{-1}AS = D \]
if and only if
\[ Av_i = \lambda_i v_i, \quad \text{all } i = 1, \ldots, n. \]

**Corollary 9.3.** \( A \in M_n(F) \) is diagonalizable over \( F \) if and only if there exists a basis of \( F^n \) consisting of eigenvectors of \( A \).

**Proof of Theorem 2.**
\[ S^{-1}AS = D \]
if and only if
\[ AS = SD \]
if and only if

\[
A[v_1 \ v_2 \ ... \ v_n] = [v_1 \ v_2 \ ... \ v_n]
\]

if and only if

\[
[Av_1 \ Av_2 \ ... \ Av_n] = [\lambda_1 v_1 \ \lambda_2 v_2 \ ... \ \lambda_n v_n]
\]

if and only if

\[
Av_i = \lambda_i v_i, \quad \text{all } i = 1, \ldots, n. \quad \square
\]

Theorems 1 and 2 provide a means of finding $S$ and $D$ to satisfy $S^{-1}AS = D$. This is referred to as diagonalizing $A$.

**Diagonalization Procedure.** Assume $A \in M_n(F)$ is given to diagonalize.

1. Find the eigenvalues of $A$, say $\lambda_1, \ldots, \lambda_n \in F$.
2. For each distinct eigenvalue of $A$, find a basis for the corresponding eigenspace by computing $\text{RREF}(A - \lambda_i I_n)$.
3. Let $B_1, \ldots, B_k$ be the bases for the $k$ distinct eigenspaces of $A$, and let $B = B_1 \cup \ldots \cup B_k$.
4. If $|B| < n$, then $A$ is not diagonalizable. If $B = \{v_1, \ldots, v_n\}$, let $S = [v_1 \ldots v_n]$ and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

(Here we must be careful to order $v_1, \ldots, v_n$ so that $Av_i = \lambda_i v_i$, all $i = 1, \ldots, n$.)

Once the procedure is understood, we have the following:

**Theorem 9.4.** For $A \in M_n(F)$, the following are equivalent:

1. $A$ is diagonalizable over $F$
2. There exists a basis of $F^n$ consisting of eigenvectors of $A$
3. All eigenvalues of $A$ are in $F$, and for each eigenvalue $\lambda_i$, we have

\[
\text{geom. mult.}(\lambda_i) = \text{alg. mult.}(\lambda_i).
\]

Another equivalent condition is contained in the next result. The proof is omitted.

**Theorem 9.5.** Let $A \in M_n(F)$. Then

$A$ is diagonalizable over $F$

if and only if there exist distinct $\lambda_1, \ldots, \lambda_k$ such that

\[
\min_A(x) = (x - \lambda_1) \ldots (x - \lambda_k).
\]
**Re-statement:** $A$ is diagonalizable over $F$ if and only if the minimal polynomial of $A$ factors into a product of distinct linear terms in $F[x]$. In the above cases $\lambda_1, \ldots \lambda_k$ are just the distinct eigenvalues of $A$. 
10. Complex Numbers

The complex numbers are
\[ \mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}, \]
where \(i^2 = -1\). The complex numbers are represented graphically by the complex plane.

If \(z = a + bi\) with \(a, b \in \mathbb{R}\), the real and imaginary parts of \(z\) are
\[ \text{Re} (z) = a \]
and
\[ \text{Im} (z) = b. \]

**Remark.**

(1) \(\text{Re} (z + w) = \text{Re} (z) + \text{Re} (w)\)
(2) \(\text{Im} (z + w) = \text{Im} (z) + \text{Im} (w)\)
(3) \(\text{Re} (zw) = \text{Re} (z)\text{Re} (w) - \text{Im} (z)\text{Im} (w)\)
(4) \(\text{Im} (zw) = \text{Im} (z)\text{Im} (w) - \text{Re} (z)\text{Re} (w)\)

The conjugate of \(z = a + bi\) is
\[ \bar{z} = a - bi. \]

**Theorem 10.1 (Properties of \(\bar{z}\)).** Let \(z, w \in \mathbb{C}\). Then

(1) \(\bar{\bar{z}} = z\)
(2) \(\bar{z} + \bar{w} = \bar{z} + \bar{w}\)
(3) \(\bar{zw} = \bar{z}\bar{w}\)
(4) \(z\) is real if and only if \(z = \bar{z}\).
If $z = a + bi$ where $a, b \in \mathbb{R}$, we define the **modulus** of $z$ as

$$|z| = \sqrt{a^2 + b^2}.$$

**Theorem 10.2** (Properties of $|z|$). Let $z, w \in \mathbb{C}$. Then

1. $|z| \geq 0$ and $|z| = 0$ if and only if $z = 0$
2. $|z|^2 = zz$
3. $|zw| = |z||w|$
4. $|z + w| \leq |z| + |w|$ and $|z + w| = |z| + |w|$ if and only if $z = 0$ or $w = \alpha z$ for $\alpha \geq 0$. (This is the **triangle inequality** for complex numbers)

We may also define the **argument** of $z \in \mathbb{C}$ if $z \neq 0$. Graphically

Here, $0 \leq \arg(z) < 2\pi$.

**Theorem 10.3** (De Moivre’s Theorem). Let $0 \neq z, w \in \mathbb{C}$. Then

$$\arg(zw) = \arg(z) + \arg(w).$$

(Actually, if $\arg(z) + \arg(w) \geq 2\pi$ we reduce by $2\pi$.)

The proof of Theorem 3 follows from elementary trigonometry.

**Theorem 10.4** (Euler’s Theorem). Let $\theta \in \mathbb{R}$. Then

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

**Proof.**

$$e^{i\theta} = 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \ldots$$

$$= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} + \ldots$$

$$= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \ldots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \ldots\right)$$

$$= \cos \theta + i \sin \theta. \quad \Box$$
Corollary 10.5 (Polar Factorization). Let $0 \neq z \in \mathbb{C}$. Then
\[ z = |z| e^{i \arg(z)}. \]

Our complete picture for $0 \neq z \in \mathbb{C}$ is

Our last result in this section is:

Theorem 10.6 (Fundamental Theorem of Algebra; C.F. Gauss). Let $p(x) \in \mathbb{C}[x]$. Then $p(x)$ factors into linear terms in $\mathbb{C}[x]$. 
11. The Adjoint of a Complex Matrix

Let \( A \in M_{m,n}(F) \). Recall that the transpose of \( A \) is
\[
A^T = [a_{ij}]^T = [a_{ji}] \in M_{n,m}(F).
\]
From math 254, we recall the following properties

1. If \( A \in M_{m,n}(F) \), then
\[
(A^T)^T = A
\]
2. If \( A, B \in M_{m,n}(F) \) and \( \alpha, \beta \in F \), then
\[
(\alpha A + \beta B)^T = \alpha A^T + \beta B^T
\]
3. If \( A \in M_{m,n}(F) \) and \( B \in M_{n,p}(F) \), then
\[
(AB)^T = B^T A^T
\]
4. If \( A \in M_n(F) \) is invertible, then
\[
(A^{-1})^T = (A^T)^{-1}
\]

If \( A \in M_{m,n}(\mathbb{C}) \) we define the conjugate of \( A \) as
\[
\bar{A} = \overline{[a_{ij}]} = [\bar{a}_{ij}].
\]
From the properties of matrix arithmetic and the properties of \( \bar{z} \), we have the following properties of \( \bar{A} \)

1. If \( A \in M_{m,n}(\mathbb{C}) \), then
\[
\overline{A} = A
\]
2. If \( A, B \in M_{m,n}(\mathbb{C}) \) and \( \alpha, \beta \in \mathbb{C} \), then
\[
\overline{\alpha A + \beta B} = \bar{\alpha} \bar{A} + \bar{\beta} \bar{B}
\]
3. If \( A \in M_{m,n}(\mathbb{C}) \) and \( B \in M_{n,p}(\mathbb{C}) \), then
\[
\overline{AB} = \bar{A} \bar{B}
\]
4. If \( A \in M_n(\mathbb{C}) \) is invertible, then
\[
\overline{(A^{-1})} = (\bar{A})^{-1}
\]
What turns out to be important for complex matrices is the adjoint or conjugate transpose of \( A \), defined by
\[
A^* = [a_{ij}]^* = [\bar{a}_{ji}] \in M_{n,m}(\mathbb{C})
\]
or
\[
A^* = (\bar{A})^T = \overline{(A^T)} \in M_{n,m}(\mathbb{C}).
\]
Properties of \( A^* \)
(1) If $A \in M_{m,n}(\mathbb{C})$, then
$$(A^*)^* = A$$

(2) If $A, B \in M_{m,n}(\mathbb{C})$ and $\alpha, \beta \in \mathbb{C}$, then
$$(\alpha A + \beta B)^* = \bar{\alpha} A^* + \bar{\beta} B^*$$

(3) If $A \in M_{m,n}(\mathbb{C})$ and $B \in M_{m,p}(\mathbb{C})$, then
$$(AB)^* = B^* A^*$$

(4) If $A \in M_n(\mathbb{C})$ is invertible, then
$$(A^*)^{-1} = (A^{-1})^*$$
12. JORDAN FORM

Suppose \( A \in M_n(\mathbb{C}) \). We know:

(1) There exist complex \( \lambda_1, \ldots, \lambda_n \in \mathbb{C} \) such that

\[
ch_A(x) = (-1)^n(x - \lambda_1) \cdots (x - \lambda_n)
\]

(This from the Fundamental Theorem of Algebra)

(2) \( A \) may or may not be diagonalizable over \( \mathbb{C} \)

(3) If \( A \) is not diagonalizable over \( \mathbb{C} \), then \( A \) is not diagonalizable over any field \( F \).

**Question.** Suppose \( A \in M_n(\mathbb{C}) \) is not diagonalizable. What “nice” matrix in \( M_n(\mathbb{C}) \) is \( A \) similar to?

The theorem concerning the **Jordan Canonical Form** will answer this question.

Let \( \lambda_0 \in \mathbb{C} \). For any \( k = 1, 2, 3, \ldots \) we define the corresponding **Jordan block matrix** of order \( k \), \( J_k(\lambda_0) \) as follows

\[
J_1(\lambda_0) = [\lambda_0]
\]

\[
J_2(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix}
\]

\[
J_3(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 \\ 0 & \lambda_0 & 1 \\ 0 & 0 & \lambda_0 \end{bmatrix}
\]

\[
J_4(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & \lambda_0 \end{bmatrix}
\]

\[
J_k(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_0 \end{bmatrix}
\]

**Example.**

(1)

\[ J_1(-3) = [-3] \]

(2)

\[ J_2(\sqrt{6}) = \begin{bmatrix} \sqrt{6} & 1 \\ 0 & \sqrt{6} \end{bmatrix} \]
Theorem 12.1. $\lambda_0$ is the only eigenvalue of $J_k(\lambda_0)$. Furthermore, $J_k(\lambda_0)$ has

1. characteristic polynomial $= (-1)^k(x - \lambda_0)^k$
2. minimal polynomial $= (x - \lambda_0)^k$
3. alg. mult.($\lambda_0$) = $k$
4. geom. mult.($\lambda_0$) = 1

We also need the notion of the **direct sum**. If $A \in M_m(F)$ and $B \in M_n(F)$, then

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in M_{m+n}(F).$$

Example.

1. $[1] \oplus \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 4 & 5 \end{bmatrix}$

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

3. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \oplus \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \oplus \begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \\ 11 & 12 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 6 & 7 & 0 \\ 0 & 0 & 0 & 0 & 8 & 9 & 10 & 0 \\ 0 & 0 & 0 & 0 & 11 & 12 & 0 & 0 \end{bmatrix}$
**Theorem 12.2.** Let \( A_i \in M_n(F) \), all \( i = 1, \ldots, t \) and let
\[
A = A_1 \oplus \cdots \oplus A_t.
\]
Then
\[
\begin{align*}
(1) & \quad A^k = A_1^k \oplus \cdots \oplus A_t^k; \text{ all } k = 1, 2, \ldots \\
(2) & \quad p(A) = p(A_1) \oplus \cdots \oplus p(A_t); \text{ all } p(x) \in F[x] \\
(3) & \quad \det(A) = \det(A_1) \cdots \det(A_t) \\
(4) & \quad \text{ch}_A(x) = \text{ch}_{A_1}(x) \cdots \text{ch}_{A_t}(x) \\
(5) & \quad \min_A(x) = \text{lcm}\{ \min_{A_1}(x), \ldots, \min_{A_t}(x) \}
\end{align*}
\]

Having defined Jordan block and direct sum, we now define that a matrix of the form
\[
J = J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k)
\]
is in **Jordan (Canonical) Form**.

**Remark.** A single Jordan block is in Jordan Form. Any diagonal matrix is in Jordan Form.

**Theorem 12.3 (Jordan Canonical Form).** Let \( A \in M_n(\mathbb{C}) \). Then \( A \) is similar to a matrix \( J \) in Jordan Form. Furthermore, \( J \) is unique except for the ordering of the Jordan blocks in \( J \).

The next theorem relates many of the quantities we have studied to the Jordan Form.

**Theorem 12.4.** Let \( A \in M_n(\mathbb{C}) \) have Jordan Form \( J \), and let \( \lambda_i \) be an eigenvalue of \( A \). Then
\[
\begin{align*}
(1) & \quad \text{The algebraic multiplicity of } \lambda_i \text{ as an eigenvalue of } A \text{ equals the sum of the sizes of the Jordan blocks in } J \text{ which correspond to } \lambda_i. \\
(2) & \quad \text{The geometric multiplicity of } \lambda_i \text{ as an eigenvalue of } A \text{ equals the number of the Jordan blocks in } J \text{ which correspond to } \lambda_i. \\
(3) & \quad \text{The multiplicity of } \lambda_i \text{ as a root of the minimal polynomial of } A \text{ equals the size of the largest Jordan block in } J \text{ which corresponds to } \lambda_i.
\end{align*}
\]
13. INNER PRODUCT SPACES AND EUCLIDEAN NORM

Let $V$ be a vector space over $F = \mathbb{R}$ or $\mathbb{C}$. An inner product on $V$ is a map

$$( , ) : V \times V \rightarrow F$$

which satisfies the following properties

1. $(v, v) \geq 0$, all $v \in V$, and $(v, v) = 0$ if and only if $v = 0$
2. $(u, v + w) = (u, v) + (u, w)$, all $u, v, w \in V$
3. $(\alpha u, v) = \alpha (u, v)$, all $u, v \in V$, all $\alpha \in F$
4. $(u, v) = (v, u)$, all $u, v \in V$.

Remark. We can use the above properties and show

$$(\alpha_1 v_1 + \alpha_2 v_2, \beta_1 u_1 + \beta_2 u_2)$$
$$= \alpha_1 \beta_1 (v_1, u_1) + \alpha_1 \beta_2 (v_1, u_2) + \alpha_2 \beta_1 (v_2, u_1) + \alpha_2 \beta_2 (v_2, u_2)$$

for all $\alpha_1, \alpha_2, \beta_1, \beta_2 \in F$ and $v_1, v_2, u_1, u_2 \in V$.

Example.

1. $V = \mathbb{R}^n$:

$$(x, y) = y^T x = \sum_{i=1}^{n} x_i y_i$$

(the standard inner product, or dot product on $\mathbb{R}^n$)

2. $V = \mathbb{C}^n$:

$$(x, y) = y^* x = \sum_{i=1}^{n} x_i \bar{y}_i$$

(the standard inner product on $\mathbb{C}^n$)

3. $V = M_{m,n}(\mathbb{C})$:

$$(A, B) = \text{trace}(B^* A) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \bar{b}_{ij}$$

(the standard inner product on $M_{m,n}(\mathbb{C})$)

4. $V = C[0, 1] = \{ f : [0, 1] \rightarrow F : f \text{ is continuous} \}$:

$$(f, g) = \int_{0}^{1} f(x) \bar{g}(x) \, dx$$

Remark. Any vector space will have an infinite number of different inner products, so the above list of examples is far from exhaustive.
Theorem 13.1 (Cauchy-Schwarz Inequality). Let \( u, v \in V \), where \( V \) is an inner product space. Then
\[
|(u, v)|^2 \leq (u, u)(v, v).
\]
Furthermore,
\[
|(u, v)|^2 = (u, u)(v, v)
\]
if and only if \( \{u, v\} \) is linearly dependent.

Proof in case \( F = \mathbb{R} \). If \( u = 0 \) or \( v = 0 \), then both sides of the inequality equal 0, and there is nothing to prove. Assume without loss of generality that \( u \neq 0 \) and \( v \neq 0 \). If \( \{u, v\} \) is linearly dependent, then \( v = \alpha u \) for some \( \alpha \in \mathbb{R} \) and
\[
|(u, v)|^2 = |(u, \alpha u)|^2 = |\alpha (u, u)|^2 = \alpha^2 (u, u)^2 = (u, u)\alpha^2 (u, u) = (u, u)(\alpha u, \alpha u) = (u, u)(v, v)
\]
so equality holds. If \( \{u, v\} \) is linearly independent, then
\[
xu + v \neq 0, \text{ all } x \in \mathbb{R}.
\]
Hence
\[
(xu + v, xu + v) > 0, \text{ all } x \in \mathbb{R}
\]
or
\[
x^2(u, u) + 2x(u, v) + (v, v) > 0, \text{ all } x \in \mathbb{R}.
\]
This implies
\[
(2(u, u))^2 < 4(u, u)(v, v)
\]
or
\[
4(u, u)^2 < 4(u, u)(v, v)
\]
or
\[
|(u, u)|^2 < (u, u)(v, v)
\]
as desired. \( \square \)

If \( v \in V \), we define the euclidean norm of \( v \) as
\[
\|v\| = \sqrt{(v, v)}.
\]

Theorem 13.2 (Properties of \( \|v\| \)). Assume \( u, v \in V \) and \( \alpha \in F \). Then
\[
(1) \|u\| \geq 0 \text{ and } \|u\| = 0 \text{ if and only if } u = 0
\]
(2) \( \|\alpha u\| = |\alpha|\|u\| \)
(3) (triangle inequality)
\[ \|u + v\| \leq \|u\| + \|v\| \]

Proof of part (3).
\[ \|u + v\| \leq \|u\| + \|v\| \]
iff
\[ \|u + v\|^2 \leq (\|u\| + \|v\|)^2 \]
iff
\[ (u, v)(u, v) \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \]
iff
\[ (u, u) + 2(u, v) + (v, v) \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \]
iff
\[ (u, v) \leq \|u\|\|v\| \]
iff
\[ |(u, v)|^2 \leq \|u\|^2\|v\|^2 \]
iff
\[ |(u, v)|^2 \leq (u, u)(v, v), \]
which is the Cauchy-Schwarz Inequality.

A vector \( u \in V \) is a **unit vector** if \( \|u\| = 1 \). Note that if \( 0 \neq v \in V \), then \( \frac{v}{\|v\|} \) is a unit vector.

Two vectors \( u, v \in V \) are **orthogonal** if and only if \( (u, v) = 0 \). More generally, \( v_1, \ldots v_k \) are orthogonal if
\[ (v_i, v_j) = 0, \text{ all } i \neq j. \]
A set of vectors \( \{v_1, \ldots v_k\} \) is **orthonormal** if and only if \( v_1, \ldots v_k \) are orthogonal unit vectors.

**Example.**
\[
S = \left\{ \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix} \right\}
\]
is orthonormal (or o.n. for short).

Suppose \( v, w \in V \) are nonzero vectors. We define the **projection of \( v \) onto \( w \)** as
\[ \text{proj}_w(v) = \text{the multiple of } w \text{ which minimizes } \|v - \alpha w\|. \]
In other words, the multiple of \( w \) which is closest to \( v \). In \( \mathbb{R}^2 \) we can draw the following picture
Remark. In the above picture, the dotted vector is \( v - \text{proj}_w(v) \). This vector is perpendicular to \( w \). In \( \mathbb{R}^2 \) (with the standard dot product) this means

\[
(v - \text{proj}_w(v), w) = 0.
\]

**Theorem 13.3.** Let \( v, w \) be nonzero vectors in an inner product space \( V \). Then

\[
\text{proj}_w(v) = \frac{(v, w)}{(w, w)} w.
\]

**Proof #1.** Let \( p = \text{proj}_w(v) \). From the previous picture, we “saw” that

\[
(v - p, w) = 0
\]

and

\[
p = \alpha w.
\]

Hence

\[
0 = (v - p, w) \\
= (v, w) - (p, w) \\
= (v, w) - (\alpha w, w) \\
= (v, w) - \alpha (w, w).
\]

Hence

\[
\alpha = \frac{(v, w)}{(w, w)}
\]

and

\[
\text{proj}_w(v) = p = \alpha w = \frac{(v, w)}{(w, w)} w. \quad \Box
\]

**Proof #2.** We find the value of \( x \) that minimizes

\[
\|v - xw\|^2 = (v - xw, v - xw) \\
= (v, v) - 2x(v, w) + x^2(w, w)
\]
by solving
\[
\frac{d}{dx} \left( (v, v) - 2x(v, w) + x^2(w, w) \right) = 0
\]
or
\[-2(v, w) + 2x(w, w) = 0.\]
Hence
\[x = \frac{(v, w)}{(w, w)}\]
and
\[
\text{proj}_w(v) = xw = \frac{(v, w)}{(w, w)} w.
\]
14. Orthogonal and Orthonormal Bases, Orthogonal Projections

Suppose \( \{v_1, \ldots, v_n\} \) is an orthogonal basis of an inner product space \( V \) and \( v \in V \). We know \( v \) can be written uniquely as a linear combination of \( v_1, \ldots, v_n \), say \( v = \alpha_1 v_1 + \cdots + \alpha_n v_n \). We can use inner products to calculate the coefficients \( \alpha_1, \ldots, \alpha_n \).

**Theorem 14.1 (Fourier Coefficients).** Let \( \{v_1, \ldots, v_n\} \) be an orthogonal basis of \( V \) and let \( v \in V \). Then

\[
v = \frac{(v, v_1)}{(v_1, v_1)} v_1 + \cdots + \frac{(v, v_n)}{(v_n, v_n)} v_n.
\]

**Proof.** Suppose \( v = \alpha_1 v_1 + \cdots + \alpha_n v_n \). Then for each \( i = 1, \ldots, n \)

\[
(v, v_i) = (\alpha_1 v_1 + \cdots + \alpha_n v_n, v_i) = \alpha_1 (v_1, v_i) + \cdots + \alpha_n (v_n, v_i) = \alpha_i (v_i, v_i).
\]

Hence \( \alpha_i = \frac{(v, v_i)}{(v_i, v_i)} \), as claimed.

**Remark.** If \( v_1, \ldots, v_n \) are orthonormal, then the formula becomes

\[
v = (v, v_1) v_1 + \cdots + (v, v_n) v_n
\]

or

\[
v = \sum_{i=1}^{n} (v, v_i) v_i.
\]

For each \( i = 1, \ldots, n \),

\[
\frac{(v, v_i)}{(v_i, v_i)} v_i = \text{proj}_{v_i} (v).
\]

Now suppose that \( W \) is a subspace of an inner product space \( V \), and that \( v \in V \). We define the **orthogonal projection of \( v \) onto \( W \)** as

\[
\text{proj}_W (v) = \text{the vector } w \in W \text{ that minimizes } \|v - w\|.
\]

This leads us to the following:

**Question.** Suppose a basis \( \{v_1, \ldots, v_n\} \) of an inner product space \( V \) is given. How can we produce an orthogonal basis of \( V \) from the given basis?

A technique which describes how to produce an orthogonal basis is as follows.
**Gram-Schmidt Orthogonalization Procedure.** Suppose linearly independent \( v_1, \ldots, v_n \) are given.

1. Let \( y_1 = v_1 \)
2. Let \( y_2 = v_2 - \frac{(v_2, y_1)}{(y_1, y_1)} y_1 \)
3. Let \( y_3 = v_3 - \frac{(v_3, y_1)}{(y_1, y_1)} y_1 - \frac{(v_3, y_2)}{(y_2, y_2)} y_2 \)

\((k + 1)\) Let \( y_{k+1} = v_{k+1} - \sum_{i=1}^{k} \frac{(v_{k+1}, y_i)}{(y_i, y_i)} y_i \)

**Theorem 14.2.** Let \( \{w_1, \ldots, w_k\} \) be an orthogonal basis of a subspace \( W \) of an inner product space \( V \), and let \( v \in V \). Then \( \text{proj}_W(v) \) exists and is unique. In fact

\[
\text{proj}_W(v) = \sum_{i=1}^{k} \frac{(v, w_i)}{(w_i, w_i)} w_i.
\]

**Remark.** We will omit the proof of the above theorem. Note that when \( V = W \), then \( \text{proj}_W(v) = v \), and Theorem 2 becomes Theorem 1.

Theorems 1 and 2 indicate the advantage of having an orthogonal basis of \( V \) or \( W \). In Theorem 1, we can use inner products to express \( v \) as a linear combination of \( v_1, \ldots, v_n \), and in Theorem 2 we can use essentially the same formula to find \( \text{proj}_W(v) \).
15. Orthogonal Subspaces and Orthogonal Complements

Let $U$ and $W$ be subspace of an inner product space $V$. We say that $U$ and $V$ are orthogonal if and only if
\[(u, w) = 0, \text{ all } u \in U, \text{ all } w \in W,\]
and write
\[U \perp W\]
to denote that $U$ is orthogonal to $W$.

Remark. If $U \perp W$, then $U \cap W = \{0\}$ since any $v \in U \cap W$ must satisfy $(v,v) = 0$.

The orthogonal complement is
\[U^\perp = \{v \in V : (v,u) = 0, \text{ all } u \in U\}.\]

Example. Let $W = \text{span}\{e_1, e_2\} \subset \mathbb{R}^3$. (i.e., $W$ is the “$x$-$y$ plane”). Then
\[W^\perp = \text{span}\{e_3\} \quad (= \text{“the } z\text{-axis”}).\]

In any inner product space $V$,
\[\{0\}^\perp = V \quad \text{and} \quad V^\perp = \{0\}.

**Theorem 15.1.** Let $U$ be a subspace of an inner product space $V$. Then

1. $U^\perp$ is a subspace of $V$,
2. $U^\perp \perp U$, and
3. $U \cap U^\perp = \{0\}$.

Remark. If $\{u_1, \ldots, u_k\}$ is a basis of $U$, then
\[v \in U^\perp \text{ if and only if } (v,u_i) = 0, \text{ all } i = 1, \ldots, k.\]

**Theorem 15.2.** Let $U$ and $W$ be subspaces of an inner product space $V$. Then
\[U \perp W \text{ if and only if } W \subset U^\perp.\]

equiv. $U \subset W^\perp$ in finite-dim case

Proof. $U \perp W$ if and only if every $v \in W$ satisfies $(v,u) = 0$ for all $u \in U$. This makes it clear from the definitions that if $U \perp W$ and $v \in W$, then $v \in U^\perp$. Hence $\subset U^\perp$. \qed

The next result gives a description of $U^\perp$ in terms of bases of $U$ and $V$. 
**Theorem 15.3.** Let $U$ be a subspace of an inner product space $V$ with $\dim(U) = k < n = \dim(V)$. Suppose $\{u_1, \ldots, u_n\}$ is an orthonormal basis of $V$ and that $\{u_1, \ldots, u_k\}$ is an orthonormal basis of $U$. Then $\{u_{k+1}, \ldots, u_n\}$ is an orthonormal basis of $U^\perp$.

**Corollary 15.4.** Let $U$ be a subspace of a finite dimensional inner product space $V$. Then

1. $(U^\perp)^\perp = U$
2. $\dim(U) + \dim(U^\perp) = \dim(V)$

The next result is from Math 254, and gives a technique for computing a basis for $U^\perp$ when $U$ is a subspace of $\mathbb{R}^n$.

**Theorem 15.5.** Let $A \in M_{m,n}(\mathbb{R})$. Then

$$\text{rowspace}(A)^\perp = \text{nullspace}(A).$$

**Remark.** Suppose $W = \text{span}\{w_1, \ldots, w_m\}$ is a subspace of $\mathbb{R}^n$, and let $A = [w_1 \ldots w_m]^T$. Then $W = \text{rowspace}(A)$ and $W^\perp = \text{nullspace}(A)$, so bases for $W$ and $W^\perp$ can be found by computing RREF$(A)$.

A similar result to Theorem 5 holds for complex matrices

**Theorem 15.6.** Let $A \in M_{m,n}(\mathbb{C})$ and let $W = \text{columnspace}(A)$, which is a subspace of $\mathbb{C}^m$. Then $W^\perp = \text{nullspace}(A)$.

Recall that if $U$ and $W$ are subspaces of $V$, then $U \cap W$ and $U + W$ are also subspaces of $V$. The next theorem describes their orthogonal complements.

**Theorem 15.7.** Let $U$ and $W$ be subspaces of a finite dimensional inner product space $V$. Then

1. $(U \cap W)^\perp = U^\perp + V^\perp$
2. $(U + W)^\perp = U^\perp \cap V^\perp$
16. SOME SPECIAL COMPLEX MATRICES

Throughout this section, all matrices will be in $M_n(\mathbb{C})$. Recall the adjoint of $A$ is

$$A^* = (\bar{A})^T,$$

and that $A^*$ satisfies

$$(Ax, y) = (x, A^*y), \text{ all } x, y \in \mathbb{C}^n.$$ 

We say that $A$ is

- **hermitian** if and only if $A^* = A$,
- **skew-hermitian** if and only if $A^* = -A$,
- **unitary** if and only if $A^* = A^{-1}$,
- **normal** if and only if $A^*A = AA^*$.

**Remark.** If $A$ is a real matrix, then these terms translate to

- **symmetric** ($A^T = A$) instead of hermitian
- **skew-symmetric** ($A^T = -A$) instead of skew-hermitian
- **orthogonal** ($A^T = A^{-1}$) instead of unitary.

**Proposition 16.1.** $A$ is hermitian if and only if $iA$ is skew-hermitian

**Proposition 16.2.** Any $A \in M_n(\mathbb{C})$ can be written uniquely as $A = H + iK$ where $H$ and $K$ are hermitian. In fact,

$$H = \frac{A + A^*}{2} = \text{Re} \ (A)$$

and

$$H = \frac{A - A^*}{2i} = \text{Im} \ (A).$$

**Theorem 16.3.** If $\lambda$ is an eigenvalue of hermitian $A$, then $\lambda$ is real.

**Proof.** Choose $x \in \mathbb{C}^n$ such that $Ax = \lambda x$ and $\|x\| = (x, x) = 1$. Then

$$\lambda = \lambda (x, x) = (\lambda x, x) = (Ax, x)$$

$$= (x, A^*x) = (x, Ax) = (Ax, x)$$

$$= (\lambda x, x) = \lambda (x, x) = \bar{\lambda}. $$
Since $\lambda = \bar{\lambda}$, $\lambda$ is real.

**Theorem 16.4.** If $\lambda$ is an eigenvalue of unitary $A$, then $|\lambda| = 1$.

**Proof.** Choose $x \in \mathbb{C}^n$ such that $Ax = \lambda x$ and $\|x\| = (x, x) = 1$. Then
\[
|\lambda|^2 = \lambda \bar{\lambda} = \lambda \bar{\lambda} (x, x) = (\lambda x, \lambda x) = (Ax, Ax) = (x, A^*Ax) = (x, x) = 1.
\]
Hence $|\lambda| = 1$. □

**Theorem 16.5.** For $A \in M_n(\mathbb{C})$, the following are equivalent (TFAE):

1. $A$ is unitary
2. $A^*A = I_n$
3. $A$ has orthonormal columns
4. $AA^* = I_n$
5. $A$ has orthonormal rows
6. $(Ax, Ay) = (x, y)$, all $x, y \in \mathbb{C}^n$
7. $(Ax, Ax) = (x, x)$, all $x \in \mathbb{C}^n$
8. $\|Ax\| = \|x\|$, all $x \in \mathbb{C}^n$

**Proof.** Note that (1) $\iff$ (2) $\iff$ (3), since for square matrices $A$ and $B$,
\[
B = A^{-1} \iff BA = I_n \iff AB = I_n.
\]
Also note that (6) $\Rightarrow$ (7) is trivial, as is (7) $\iff$ (8).

Since for any $A$, $A^*A = [\langle \text{col}(A), \text{col}(A) \rangle]$, we see that (2) $\iff$ (3). In a similar way, (4) $\iff$ (5).

Note that (1) $\Rightarrow$ (6), since if $A^* = A^{-1}$, then
\[
(Ax, Ay) = (x, A^*Ay) = (x, A^{-1}Ay) = (x, y).
\]
So far we have shown (1) $\iff$ (2) $\iff$ (3) $\iff$ (4) $\iff$ (5), (1) $\Rightarrow$ (6), (6) $\Rightarrow$ (7), and (7) $\iff$ (8). It remains to show (7) $\Rightarrow$ (6) and (6) $\Rightarrow$ (1).

To show (7) $\Rightarrow$ (6), suppose $x, y \in \mathbb{C}^n$ are given. Expanding
\[
(A(x + y), A(x + y)) = (x + y, x + y)
\]
gives
\[
(Ax, Ax) + (Ax, Ay) + (Ay, Ax) + (Ay, Ay) = (x, x) + (x, y) + (y, x) + (y, y).
\]
Hence
\[
(Ax, Ay) + (Ay, Ax) = (x, y) + (y, x).
\]
(\dag)

Applying the same trick to
\[
(Ax + iy, Ax + iy)) = (x + iy, x + iy)
\]
yields
\[(Ax, Ai y) + (Ai y, Ax) = (x, iy) + (iy, x)\]
or
\[-i(Ax, Ay) + i(Ay, Ax) = -i(x, y) + i(y, x)\]
or
\[(‡) (Ax, Ay) - (Ay, Ax) = (x, y) - (y, x).\]
Adding (†) and (‡) gives
\[2(Ax, Ay) = 2(x, y),\]
and so
\[(Ax, Ay) = (x, y),\]
and we have shown that (7) ⇒ (6).
To complete the proof, it suffices to show (6) ⇒ (1). If
\[(Ax, Ay) = (x, y), \text{ all } x, y \in \mathbb{C}^n,\]
then
\[(A^*Ax, y) = (x, y), \text{ all } x, y \in \mathbb{C}^n\]
\[(A^*Ax, y) - (x, y) = 0, \text{ all } x, y \in \mathbb{C}^n\]
\[(A^*Ax - x, y) = 0, \text{ all } x, y \in \mathbb{C}^n\]
\[((A^*A - I_n)x, y) = 0, \text{ all } x, y \in \mathbb{C}^n.\]
This last can only happen in case \(A^*A - I_n = 0\), which implies \(A^* = A^{-1}.\)

**Theorem 16.6.** Let \(A, U \in M_n(\mathbb{C})\) with \(U\) unitary. Then
\(A\) is normal if and only if \(U^*AU\) is normal.

**Proof.** Assume \(A\) is normal. Then
\[
(U^*AU)^*U^*AU = U^*A^*UU^*AU
= U^*A^*I_nAU
= U^*A^*AU
= U^*AA^*U
= U^*AUU^*A^*U
= U^*AU(U^*AU)^*.
\]
Hence, \(U^*AU\) is normal.
Since \(U^*\) is also unitary, the previous argument can be used to show
\(U^*AU\) is normal ⇒ \(U(U^*AU)U^*\) is normal ⇒ \(A\) is normal. \(\square\)
17. **Unitary Similarity and Schur’s Theorem**

Let \( A, B \in M_n(\mathbb{C}) \). We say that \( A \) and \( B \) are **unitarily similar** if

\[
U^* A U = B
\]

for some unitary \( U \in M_n(\mathbb{C}) \).

*Remark.* If \( A \) and \( B \) are unitarily similar, then \( A \) and \( B \) are similar, since \( U^* = U^{-1} \). The converse is false.

*Example.* Let

\[
A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.
\]

Then \( A \) and \( B \) are similar. However, we will see that \( A \) and \( B \) are not unitarily similar.

**Theorem 17.1.** *Unitarily similarity is an equivalence relation.* In other words:

(1) \( A \) is unitarily similar to \( A \) for all \( A \in M_n(\mathbb{C}) \),

(2) \( A \) is unitarily similar to \( B \) if and only if \( B \) is unitarily similar to \( A \),

(3) If \( A \) is unitarily similar to \( B \), and \( B \) is unitarily similar to \( C \), then \( A \) is unitarily similar to \( C \).

*Question.*

(1) When is \( A \) unitarily similar to a diagonal matrix? (This will be answered in the next section.)

(2) If \( A \) is not unitarily similar to a diagonal matrix, then what “nice” matrix is \( A \) unitarily similar to? (The answer is contained in the next Theorem.)

**Theorem 17.2.** Let \( A \in M_n(\mathbb{C}) \) have eigenvalues \( \lambda_1, \ldots, \lambda_n \) (in any order). Then there exists unitary \( U \) and upper triangular \( T \) in \( M_n(\mathbb{C}) \) such that

\[
U^* A U = T
\]

and

\[
diag(T) = (\lambda_1, \ldots, \lambda_n).
\]

That is:

\[
T = \begin{bmatrix}
\lambda_1 & ? & \cdots & ? \\
0 & \lambda_2 & \cdots & ? \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \lambda_{n-1} & ? \\
0 & 0 & 0 & \cdots & \lambda_n
\end{bmatrix}.
\]
Lemma 17.3. Let $A \in M_n(F)$ and suppose $S \in M_n(F)$ is invertible. Then

$$S^{-1}AS = \begin{bmatrix} \lambda_1 & ? & \cdots & ? \\ 0 & \vdots & & A_1 \\ \vdots & \ddots & & \vdots \\ 0 & & \ddots & \vdots \\ \end{bmatrix},$$

where $A_1 \in M_{n-1}(F)$ if and only if

$$Av = \lambda_1 v,$$

where $v = \text{col}_1(S)$.

Proof of Lemma 3.

$$S^{-1}AS = \begin{bmatrix} \lambda_1 & ? & \cdots & ? \\ 0 & \vdots & & A_1 \\ \vdots & \ddots & & \vdots \\ 0 & & \ddots & \vdots \\ \end{bmatrix}$$

if and only if

$$AS = S \begin{bmatrix} \lambda_1 & ? & \cdots & ? \\ 0 & \vdots & & A_1 \\ \vdots & \ddots & & \vdots \\ 0 & & \ddots & \vdots \\ \end{bmatrix}$$

if and only if

$$A[v_1 \ldots v_n] = [v_1 \ldots v_n] \begin{bmatrix} \lambda_1 & ? & \cdots & ? \\ 0 & \vdots & & \ast \\ \vdots & \ddots & & 0 \\ \end{bmatrix}$$

if and only if

$$Av = \lambda_1 v.$$  \hfill $\square$

The proof of Schur’s Theorem relies on the above lemma and uses induction on $n$. As in the proof of Lemma 3, matrix multiplication plays a crucial role.

Proof of Schur’s Theorem. Assume that Schur’s Theorem holds for matrices in $M_{n-1}(\mathbb{C})$. That is, if $A_1 \in M_{n-1}(\mathbb{C})$ has eigenvalues $\lambda_2, \ldots, \lambda_n$, then there exists unitary $U_1 \in M_{n-1}(\mathbb{C})$ and upper triangular $T \in M_{n-1}(\mathbb{C})$ such that

$$U_1^* A_1 U_1 = T_1$$
and

\[
T_1 = \begin{bmatrix}
\lambda_2 & \cdots & ? \\
0 & \lambda_3 & \cdots & ? \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{bmatrix}.
\]

Choose an eigenvector \( v_1 \in \mathbb{C}^n \) such that \( Av_1 = \lambda_1 v_1 \) and \( \|v_1\| = 1 \). Extend \( v_1 \) to an orthonormal basis of \( \mathbb{C}^n \), say \( v_1, \ldots, v_n \). Let

\[
V = [v_1 \ldots v_n] \in M_n(\mathbb{C}).
\]

Since \( V \) has orthonormal columns, \( V \) is unitary. By Lemma 3,

\[
V^*AV = \begin{bmatrix}
\lambda_1 & \cdots & ? \\
0 & \ddots & \vdots \\
\vdots & \ddots & A_1 \\
0 & \cdots & \lambda_n
\end{bmatrix}
\]

where \( A_1 \in M_{n-1}(\mathbb{C}) \) has eigenvalues \( \lambda_2, \ldots, \lambda_n \).

By the induction hypothesis, there exists unitary \( U_1 \in M_{n-1}(\mathbb{C}) \) such that

\[
U_1^* A_1 U_1 = T_1 = \begin{bmatrix}
\lambda_2 & \cdots & ? \\
0 & \lambda_3 & \cdots & ? \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & \lambda_n
\end{bmatrix}.
\]

Let \( W = 1 \oplus U_1 \in M_n(\mathbb{C}) \). Since \( W \) has orthonormal columns, \( W \) is unitary.

\[
W = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & \vdots \\
\vdots & \ddots & \ddots & U_1 \\
0 & \cdots & 0 & 0
\end{bmatrix}
\]
Let \( U = VW \). Since \( V \) and \( W \) are unitary, \( U \) is also unitary. We calculate

\[
U^*AU = (VW)^*AVW
\]

\[
= W^* \begin{bmatrix}
\lambda_1 & ? & \ldots & ? \\
0 & & & \\
\vdots & & & A_1 \\
0 & & & \\
\end{bmatrix} W
\]

\[
= \begin{bmatrix}
1 & 0 & \ldots & 0 \\
0 & & & \\
\vdots & U_1^* & & A_1 \\
0 & & & \\
\end{bmatrix} \begin{bmatrix}
\lambda_1 & ? & \ldots & ? \\
0 & & & \\
\vdots & & & U_1 \\
0 & & & \\
\end{bmatrix} \begin{bmatrix}
1 & 0 & \ldots & 0 \\
0 & & & \\
\vdots & U_1^*A_1 & & U_1 \\
0 & & & \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\lambda_1 & ? & \ldots & ? \\
0 & & & \\
\vdots & U_1^*A_1U_1 & & \\
0 & & & \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\lambda_1 & ? & \ldots & ? \\
0 & & & \\
\vdots & T_1 & & \\
0 & & & \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\lambda_1 & ? & \ldots & ? \\
0 & \lambda_2 & ? & \ldots & ? \\
0 & 0 & \lambda_3 & \ddots & ? \\
\vdots & \ddots & & \ddots & \ddots \\
0 & 0 & \ldots & 0 & \lambda_n
\end{bmatrix} = T. \quad \square
\]
18. Unitary Diagonalization and the Spectral Theorem

We say that $A \in M_n(\mathbb{C})$ is unitarily diagonalizable if and only if there exists unitary $D$ and diagonal $D$ such that

$$U^*AU = D.$$ 

Remark. If $A$ is unitarily diagonalizable, then $A$ is diagonalizable, since $U^* = U^{-1}$.

Recall that if $S, A, D \in M_n(F)$ with $S$ invertible and $D$ diagonal, then

$$S^{-1}AS = D$$

if and only if the columns of $S$ are a basis of $F^n$ consisting of eigenvectors of $A$. Furthermore, the eigenvalue associated with the $i$th column of $S$ is the $i$th diagonal entry of $A$.

The next theorem provides a first answer to the question “When is $A$ unitarily diagonalizable?”.

**Theorem 18.1.** $A \in M_n(\mathbb{C})$ is unitarily diagonalizable if and only if there exists an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $A$.

**Proof.** ($\Rightarrow$) Suppose $U^*AU = D$ where $U = [u_1 \ u_2 \ldots \ u_n]$ is unitary and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Then we have the following sequence of equivalent statements

$$U^*AU = D$$

$$AU = UD$$

$$A[u_1 \ u_2 \ldots \ u_n] = [u_1 \ u_2 \ldots \ u_n] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}$$

$$A[u_1 \ u_2 \ldots \ u_n] = [\lambda_1u_1 \ \lambda_2u_2 \ldots \ \lambda_nu_n]$$

$$Au_i = \lambda_i u_i, \ \text{all } i = 1, \ldots, n.$$ 

Hence the columns of $U$ are eigenvectors of $A$. Furthermore, $u_1, \ldots, u_n$ are orthonormal since $U$ is unitary.

($\Leftarrow$) Suppose $\{u_1, \ldots, u_n\}$ is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $A$. Say

$$Au_i = \lambda_i u_i, \ \text{all } i = 1, \ldots, n.$$
Let $U = [u_1 \ldots u_n]$ and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Since $u_1, \ldots, u_n$ are orthonormal, $A$ is unitary. Furthermore:

$$AU = A[u_1 \ldots u_n] = [Au_1 \ldots Au_n] = [\lambda_1 u_1 \ldots \lambda_n u_n] = UD.$$ 

Hence $U^*AU = U^*(UD) = D$, and $A$ is unitarily diagonalizable. □

Before we give a second answer to the question of when $A$ is unitarily diagonalizable, let us state some lemmas.

**Lemma 18.2.** If $A, U \in M_n(\mathbb{C})$ with $U$ unitary, then $A$ is normal if and only if $U^*AU$ is normal.

(This is Theorem 14.4)

**Lemma 18.3.** If $A \in M_n(\mathbb{C})$, then there exists unitary $U$ and upper triangular $T$ such that $U^*AU = T$.

(This is Schur’s Theorem, Theorem 15.2)

**Lemma 18.4.** Let $T \in M_n(\mathbb{C})$ be upper triangular. Then $T$ is normal if and only if $T$ is diagonal.

We defer the proof of lemma 4, which relies on matrix multiplication and an examination of the equation $T^*T = TT^*$.

We may now state the main result.

**Theorem 18.5** (Spectral Theorem for Normal Matrices). Let $A \in M_n(\mathbb{C})$. Then $A$ is unitarily diagonalizable if and only if $A$ is normal.

**Proof.** ($\Rightarrow$) Assume $U^*AU = D$, where $U$ is unitary and $D$ is diagonal. Since $D$ is diagonal it is normal. By lemma 2, $A$ is normal.

($\Leftarrow$) Assume $A$ is normal. By lemma 3, there exists unitary $U$ and upper triangular $T$ such that $U^*AU = T$. By lemma 2 and $A$ being normal, $T$ is normal. Since $T$ is normal, lemma 4 yields that $T$ is diagonal, and so $A$ is unitarily diagonalizable. □

To complete the proof of the Spectral Theorem, we need to prove lemma 4. For the proof we need the facts that

$$||Ae_i|| = ||\text{col}_i(A)||$$

$$||A^*e_i|| = ||\text{row}_i(A)||.$$ 

**Proof of lemma 4.** ($\Leftarrow$) If $T$ is diagonal, then $T$ is obviously normal.
(⇒) Let
\[
T = \begin{bmatrix}
t_{11} & t_{12} & \cdots & t_{1n} \\
0 & t_{22} & \cdots & t_{2n} \\
0 & \cdots & \ddots & \vdots \\
0 & \cdots & 0 & t_{nn}
\end{bmatrix}
\]
and assume \(T^*T = TT^*\). First, consider that
\[
(T^*T)_{1,1} = (TT^*)_{1,1},
\]
\[
(T^*T)_{1,1} = |t_{11}|^2,
\]
and \((TT^*)_{1,1} = |t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2\).
Hence \(t_{12} = \cdots = t_{1n} = 0\) and
\[
T = \begin{bmatrix}
t_{11} & 0 & \cdots & 0 \\
0 & t_{22} & \cdots & t_{2n} \\
0 & \cdots & \ddots & \vdots \\
0 & \cdots & 0 & t_{nn}
\end{bmatrix}.
\]

Next consider
\[
(T^*T)_{2,2} = (TT^*)_{2,2},
\]
\[
(T^*T)_{2,2} = |t_{22}|^2,
\]
and \((TT^*)_{2,2} = |t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2\).
Hence \(t_{23} = \cdots = t_{2n} = 0\) and
\[
T = \begin{bmatrix}
t_{11} & 0 & \cdots & 0 \\
0 & t_{22} & \cdots & 0 \\
0 & \cdots & \ddots & \vdots \\
0 & \cdots & 0 & t_{nn}
\end{bmatrix}.
\]

We proceed in this fashion, using
\[
(T^*T)_{i,i} = (TT^*)_{i,i}, \text{ all } i = 3, 4, \ldots, n
\]
to obtain
\[
t_{ij} = 0, \text{ all } i < j.
\]
Hence \(T\) is diagonal. \(\square\)

The next theorem concerns hermitian matrices. The proof illustrates how the Spectral Theorem can be used in proofs.

**Theorem 18.6.** Let \(A \in M_n(\mathbb{C})\). Then \(A\) is normal with real eigenvalues if and only if \(A\) is hermitian.
Proof. \((\Rightarrow)\) Assume \(A\) is normal with real eigenvalues. Since \(A\) is normal, the Spectral Theorem implies that \(U^*AU = D\) for some unitary \(U\) and diagonal \(D\). Since \(A\) has real eigenvalues, \(D\) is real. Then \(A = UDU^*\) and

\[ A^* = (UDU^*)^* = U^{**}D^*U^* = UDU^* = A. \]

Hence \(A\) is hermitian.

\((\Leftarrow)\) Assume \(A^* = A\). Then \(A\) is normal, since \(A^*A = A^2 = AA^*.\) By theorem 14.1, all the eigenvalues of \(A\) are real. \(\square\)
19. **UNITARY EQUIVALENCE AND THE SINGULAR VALUE DECOMPOSITION**

**Definition.** We say that \( A, B \in M_n(\mathbb{C}) \) are *unitarily equivalent* if and only if there exist unitary \( U, V \in M_n(\mathbb{C}) \) such that

\[
UAV = B.
\]

**Theorem 19.1.** Unitary equivalence is an equivalence relation on \( M_n(\mathbb{C}) \).

**Question.** To what “nice” matrix is \( A \in M_n(\mathbb{C}) \) unitarily equivalent?

**Theorem 19.2** (Singular Value Decomposition). Let \( A \in M_n(\mathbb{C}) \). There exist unitary \( U, V \in M_n(\mathbb{C}) \) such that

\[
UAV = S = \text{diag}(s_1, \ldots, s_n)
\]

where \( s_1 \geq \cdots \geq s_n \geq 0 \) are the eigenvalues of \( A^*A \) and/or \( AA^* \), and \( s_1, \ldots, s_n \) are referred to as the *singular values* of \( A \).

We will not prove theorem 1, but will use theorem 1 to define two related items in matrix theory. These are the generalized inverse and the polar factorization.

If \( \text{rank}(A) = k \), then \( s_1 \geq \cdots \geq s_k = s_{k+1} = \cdots = s_n \) and the matrix \( S \) in theorem 1 is of the form

\[
S = \\
\begin{bmatrix}
    s_1 & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\
    0 & s_2 & \ldots & \ldots & \ldots & \ldots & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
    \vdots & \vdots & \ldots & s_k & 0 & \ldots & 0 \\
    \vdots & \vdots & \ldots & \vdots & \vdots & \ddots & \vdots \\
    \vdots & \vdots & \ldots & \vdots & \vdots & \ldots & 0 \\
    0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

We define

\[
S^\# = \\
\begin{bmatrix}
    s_1^{-1} & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\
    0 & s_2^{-1} & \ldots & \ldots & \ldots & \ldots & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
    \vdots & \vdots & \ldots & s_k^{-1} & 0 & \ldots & 0 \\
    \vdots & \vdots & \ldots & \vdots & \vdots & \ddots & \vdots \\
    \vdots & \vdots & \ldots & \vdots & \vdots & \ldots & 0 \\
    0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]
Definition. If $UAV = S$, then $A = U^*SV^*$, and we define the\ generalized inverse, or pseudo-inverse or Morse-Penrose inverse of $A$ to be

$$A^\# = (U^*SV^*)^\# = V S^\# U.$$ 

Remark. It is possible to state theorem 2 and define the generalized inverse for $A \in M_{m,n}(\mathbb{C})$. Some properties of $A^\#$ are contained in the following result.

**Theorem 19.3.** Let $A \in M_n(\mathbb{C})$. Then

1. $(AA^\#)^* = AA^\#
2. (A^\#A)^* = A^\#A
3. (AA^\#)^2 = AA^\#
4. (A^\#A)^2 = A^\#A
5. AA^\#A = A^\#
6. A^\#AA^\# = A^\#
7. (A^\#)^* = A^\#
8. If $A$ is invertible, then $A^\# = A^{-1}$
9. $x = A^\#b$ is the vector of minimum norm which minimizes $\|Ax - b\|^2$.

We will consider property 9 again later.

**Theorem 19.4 (Polar factorization).** Let $A \in M_n(\mathbb{C})$. Then $A$ can be factored as $A = KW$ and $A = WH$ where $W$ is unitary and $H, K$ are hermitian, with the eigenvalues of $H, K$ equal to $s_1, \ldots, s_n$ the singular values of $A$.

**Proof sketch.** Suppose $UAV = S$ is the singular value decomposition of $A$. Define

$$W = U^*V^*, \quad H = VSV^*, \quad K = U^*SU.$$ 

It can then be shown that $W, H, K$ have the desired properties. □

Remark. If $n = 1$, the Polar Factorization of $a \in \mathbb{C}$ is

$$a = |a|e^{i\text{arg}a},$$

where $|a|$ plays the role of $H, K$, and $e^{i\text{arg}a}$ plays the role of $W$. 

adjoint, 34, 48
basis, 5
Cayley-Hamilton Theorem, 24
complex numbers, 31
coordinate vector, 11
diagonalizable, 27
dimension, 6
direct sum, 37
eigen-, 14	space, 14
value, 14
vector, 14
euclidean norm, 40
hermitian, 48
skew-, 48
inner product, 39
orthogonal, 41, 46
complement, 46
orthonormal, 41
Jordan
block matrix, 36
Canonical Form, 38
linear transformation, 7
linearly independent, 5
multiplicity, 15
algebraic, 20
geometric, 15
normal, 48
nullity, 6, 8
nullspace, kernel, 8
orthogonal matrix, 48
polynomial
annihilating, 22
characteristic, 17
minimum, 23
projection, 41
orthogonal, 44
rank, 6, 8
rowspace, 6
similarity, 25
shared properties, 25
unitary, 51
span, 4
Spectral Theorem, 56
submatrix, 18
principal, 19
subdeterminant, 18
principal subdeterminant, 19
Sylvester's Law of Nullity, 6
symmetric, 48
skew-, 48
transpose, 34
unitarily diagonalizable, 55
unitary, 48
vector space, 3
examples, 3
SAN DIEGO STATE UNIVERSITY
E-mail address: grone@math.sdsu.edu

UNIVERSITY OF CALIFORNIA, RIVERSIDE
E-mail address: jmd@math.ucr.edu