University of California, Berkeley

Literally everything I know about

Linear Algebra

Warren Alphonso

A very reductionist summary of Linear Algebra and its Applications by Lay, Lay, and McDonald, Linear Algebra Done Wrong by Treil, and Quantum Computation and Quantum Information by Nielsen and Chuang.

July 26, 2019

Contents

Chapter 1

Basic Notations

1.1 Vector Spaces

Definition 1.1.1. A **vector space** V is a collection of vectors, along with vector addition and scalar multiplication defined such that for vectors u, v, and w:

1. Commutative: $v + w = w + v$
2. Associative: $(u + v) + w = u + (v + w)$
3. Zero vector: $v + 0 = v$
4. Additive inverse: $v + (-v) = 0$
5. Multiplicative identity: $1v = v$
6. Multiplicative associative: $(\alpha\beta)v = \alpha(\beta v)$
7. Distribution of scalars: $\alpha(u + v) = \alpha u + \alpha v$
8. Distribution of vectors: $(\alpha + \beta)u = \alpha u + \beta u$

These properties ensure that vector spaces are **abelian groups**.

Definition 1.1.2. An $m \times n$ **matrix** is an array with m rows and n columns. Elements of a matrix are called *entries*. Given a matrix A, its **transpose** is defined as the matrix whose columns are A's rows, so A^T is an $n \times m$ matrix.

1.2 Linear Combinations

Definition 1.2.1. A **linear combination** of vectors $v_1, \cdots, v_p \in V$ is a sum of the form

$$\alpha_1 v_1 + \cdots + \alpha_p v_p = \sum_{k=1}^{p} \alpha_k v_k$$

Definition 1.2.2. A set of vectors $v_1, \cdots, v_n$ is said to be **linearly independent** if the equation

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$$

has only the trivial solution where all coefficients are 0.

Definition 1.2.3. A **basis** is a set of vectors $v_1, \cdots, v_n \in V$ such that any vector $u \in V$ has a *unique* representation as a linear combination

$$u = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

The coefficients $\alpha_1, \cdots, \alpha_n$ are called *coordinates* of u.

Fundamentally, our definition of basis requires that it must be spanning and unique. In order for a representation to be unique, we know the basis must be linearly independent.

Theorem 1.2.1. *A set of vectors $v_1, \cdots, v_p \in V$ is a basis if and only if it is linearly independent and complete (spanning).*

Proof. We already know a basis must be linearly independent and spanning, so we just need to prove the other direction.
Suppose the set $v_1, \cdots, v_p$ is linearly independent. Then we know for some vector $u \in V$:

$$u = \sum_{k=1}^{n} \alpha_k v_k$$

All that is remaining is to prove this representation is unique.
Suppose there is another representation, $u = \sum_{k=1}^{n} \beta_k v_k$. Then,

$$\sum_{k=1}^{n} (\alpha_k - \beta_k) v_k = v - v = 0$$

Since the set is linearly independent, we know $\alpha_k - \beta_k = 0$. Thus, the representation is unique. ■

1.3 Linear Transformations

Definition 1.3.1. A **transformation** T from set X to set Y assigns a value $y \in Y$ for every value $x \in X$: $y = T(x)$. X is called the *domain* of T, Y is called the *codomain* of T, and the set of all $T(x)$ is called the *range* of T.
Let V, W be vector spaces. A transformation $T : V \to W$ is **linear** if:

1. $T(u + v) = T(u) + T(v)$
2. $T(\alpha v) = \alpha T(v)$

A mapping $T : \mathbb{F}^n \to \mathbb{F}^m$ is *onto* $\mathbb{F}^m$ if each b in $\mathbb{F}^m$ is the image of at least one x in $\mathbb{F}^n$.
A mapping $T : \mathbb{F}^n \to \mathbb{F}^m$ is *one-to-one* if each b in $\mathbb{F}^m$ is the image of at most one x in $\mathbb{F}^n$.

We can represent linear transformations with matrices. To represent a transformation $T : \mathbb{F}^n \to \mathbb{F}^m$, we need to only know our n basis vectors are transformed. To see this, note that any vector $u = \alpha_1 v_1 + \cdots + \alpha_n v_n$. So $T(u) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$. If we join the vectors $T(v_1), \cdots, T(v_n)$ in a matrix $A = \begin{bmatrix} T(v_1) & \cdots & T(v_n) \end{bmatrix}$, we have captured all the information about T.

Definition 1.3.2. There are two ways to approach **matrix-vector multiplication**:
Column by coordinate rule: Multiply each column of the matrix by the corresponding coordinate of the

vector and add.

$$\begin{bmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = 1\begin{bmatrix}1\\3\end{bmatrix} + 2\begin{bmatrix}2\\2\end{bmatrix} + 3\begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}14\\10\end{bmatrix}$$

Row by column rule: To get entry k of the result, multiply row k of the matrix with the vector.

$$\begin{bmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}1\cdot 1 + 2\cdot 2 + 3\cdot 3\\ 3\cdot 1 + 2\cdot 2 + 1\cdot 3\end{bmatrix} = \begin{bmatrix}14\\10\end{bmatrix}$$

Definition 1.3.3. The natural extension to **matrix multiplication** of two matrices AB is to multiply A by each column of B.

$$AB = A\begin{bmatrix}b_1 & \cdots & b_n\end{bmatrix} = \begin{bmatrix}Ab_1 & \cdots & Ab_n\end{bmatrix}$$

Using the *row by columns rule*, we can see that the $(AB)_{j,k} = (\text{row } j \text{ of } A)\cdot(\text{column } k \text{ of } B)$. This also means AB is only defined if A is $m \times n$ and B is $n \times r$.
Let A be an $m \times n$ matrix and let B and C have sizes for which the indicated sums and products are defined. Then:

1. $A(BC) = (AB)C$
2. $A(B + C) = AB + AC$
3. $(B + C)A = BA + CA$
4. $\alpha(AB) = (\alpha A)B = A(\alpha B)$
5. $I_m A = A = AI_n$

Warnings:

1. In general, $AB \neq BA$
2. If $AB = AC$, then it is **not true** in general that $B = C$
3. If $AB = 0$, then it is **not true** always that $A = 0$ or $B = 0$

Definition 1.3.4. The **transpose** of A is the matrix whose columns are formed from the corresponding rows of A, denoted by A^T.

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(cA)^T = cA^T$
4. $(AB)^T = B^T A^T$

To understand the final property, let AB denote a $n \times m$ matrix so that

$$AB = \begin{bmatrix} A_{1*} \cdot B_1 & A_{1*} \cdot B_2 & \cdots & A_{1*} \cdot B_m \\ A_{2*} \cdot B_1 & A_{2*} \cdot B_2 & \cdots & A_{2*} \cdot B_m \\ \vdots & \vdots & & \vdots \\ A_{n*} \cdot B_1 & A_{n*} \cdot B_2 & \cdots & A_{n*} \cdot B_m \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} A_{1*} \cdot B_1 & A_{2*} \cdot B_1 & \cdots & A_{n*} \cdot B_1 \\ A_{1*} \cdot B_2 & A_{2*} \cdot B_2 & \cdots & A_{n*} \cdot B_2 \\ \vdots & \vdots & & \vdots \\ A_{1*} \cdot B_m & A_{2*} \cdot B_m & \cdots & A_{n*} \cdot B_m \end{bmatrix}$$

where $A_{\alpha *}$ denotes the αth row of a A. Note that $(AB)^T_{jk} = (\text{row } k \text{ of } A) \cdot (\text{column } j \text{ of } B) = (\text{row } j \text{ of } B^T) \cdot (\text{column } k \text{ of } A^T)$.

Definition 1.3.5. For a *square* matrix A, its **trace** is the sum of its diagonal entries.

$$trace(A) = \sum_{k=1}^{n} a_{k,k}$$

Theorem 1.3.1. *Let A and B be sizes $m \times n$ and $n \times m$, respectively. Then*

$$trace(AB) = trace(BA)$$

Proof. We need only show that the diagonal entries of AB are the same as the diagonal entries of BA.

$$tr(AB) = \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij} = \sum_{j=1}^{n} (BA)_{jj} = tr(BA)$$

■

1.3.1 Linear Transformations as a Vector Space

Let's further abstract the notion of a linear transformation by considering the collection of *all* linear transformations from V to W, denoted $L(V, W)$.

For any linear transformation $T \in L(V, W)$, we can define a new transformation αT. We can prove this transformation is linear:

$$\begin{aligned} (\alpha T)(\alpha_1 v_1 + \alpha_2 v_2) &= \alpha(T(\alpha_1 v_1 + \alpha_2 v_2)) \\ &= \alpha(\alpha_1 T v_1 + \alpha_2 T v_2) \\ &= \alpha_1 (\alpha T) v_1 + \alpha_2 (\alpha T) v_2 \end{aligned}$$

where the second step follows from the linearity of T.

A similar proof can be made to show that the sum of any $T_1, T_2 \in L(V, W)$ is also linear, which means it's in the set $L(V, W)$. This means we have defined multiplication by a scalar and addition on $L(V, W)$, which means $L(V, W)$ is a vector space.

1.4 Invertible Transformations and Isomorphisms

Definition 1.4.1. An $n \times n$ matrix A is **invertible** if there is an $n \times n$ matrix A^{-1} such that $A^{-1}A = I$ and $AA^{-1} = I$.
An $n \times n$ matrix A is *left invertible* if there is matrix B such that $BA = I$ and is *right invertible* if there is a matrix C such that $AC = I$. If A is both left and right invertible, then A is called **invertible**.

Theorem 1.4.1. *If A and B are invertible (and AB is defined), then the product AB is invertible and*

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof. Direct computation:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$$

and

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$$

■

Theorem 1.4.2. *If A is invertible, then A^T is also invertible and*

$$(A^{-1})^T = (A^T)^{-1}$$

Proof. Using $(AB)^T = B^T A^T$,

$$(A^{-1})^T(A^T) = (AA^{-1})^T = I$$

and

$$A^T(A^{-1})^T = (A^{-1}A)^T = I$$

■

Definition 1.4.2. An invertible linear transformation $A : V \to W$ is called an **isomorphism**. The two vector spaces V and W for which A is defined are called **isomorphic**, denoted $V \cong W$.

Isomorphic spaces can be understood as different representations of the *same* space. To see this,

Theorem 1.4.3. *Let $A : V \to W$ be an isomorphism, and let $v_1, \cdots, v_n$ be a basis in V. Then $Av_1, \cdots, Av_n$ is a basis in W.*

Proof. Because V and W are isomorphic, every $w \in W$ can be represented as some $v \in V$ by applying A^{-1}. For arbitrary $w \in W$

$$A^{-1}w = v = \sum_{k=1}^{n} \alpha_k v_k$$

Then we apply A to get

$$Av = w = \sum_{k=1}^{n} \alpha_k Av_k$$

■

Chapter 2

Systems of Linear Equations

2.1 Representations of Linear Systems

The first understanding of a *linear system* is simply a collection of m linear equations with n unknowns $x_1, \cdots, x_n$. To solve this system entails finding all n-tuples of numbers $x_1, \cdots, x_n$ which satisfy the m equations simultaneously. If we define

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

then we can summarize our linear system in matrix form

$$Ax = b$$

The above is the **coefficient matrix**. If we want to contain all the information in a single matrix, we can use an **augmented matrix**

$$\left[\begin{array}{cccc|c} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} & b_1 \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} & b_m \end{array}\right]$$

2.2 Solving Linear Systems

Linear systems are solved using **Gaussian elimination**. We can perform the following row operation on an augmented matrix:

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

These operations belong to the *elementary matrices*: any operation can be described by applying the same operation to I to get E and then multiplying EA.

Definition 2.2.1. For an augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 2 & 1 & 7 \\ 2 & 1 & 2 & 1 \end{array}\right]$$

the **echelon form** is

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 2 & -4 \end{array}\right]$$

and the **reduced echelon form** is

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

Formally, the **echelon form** requires that all zero rows are below all nonzero rows and that any nonzero row's **pivot**, its leading entry, is strictly to the right of the leading entry in the previous row. The particular echelon form above is called **triangular form** and is only possible when we have a square matrix. The **reduced echelon form** requires echelon form in addition to maintaining that all pivot entries are 1 and that all entries above each pivot are 0.

The *existence* and *uniqueness* of a solution can be determined by analyzing pivots in the echelon form of a matrix.
When looking at the coefficient matrix:

1. A solution (if it exists) is unique if and only if there are no free variables, that is if the echelon form has a pivot in every *column*.
2. A solution is consistent if and only if the echelon form has a pivot in every *row*.

The first statement is trivial because free variables are responsible for all non-uniqueness. For the second statement, if we have a row with no pivots in the echelon form of a matrix, we have $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b_k \end{array}\right]$, which certainly has no solution. Thus, in order for a solution to *exist* and be *unique*, the echelon form must have a pivot in *every column and every row*.

Theorem 2.2.1. *Any linearly independent system of vectors in $\mathbb{F}^n$ cannot have more than n vectors in it.*

Proof. Let a system $v_1, \cdots, v_m \in \mathbb{F}^n$ be linearly independent and let $A = \left[\begin{array}{ccc} v_1 & \cdots & v_m \end{array}\right]$ be $n \times m$. We must show that $x_1 v_1 + \cdots + x_m v_m = 0$, or equivalently $Ax = 0$, has unique solution $x = 0$. According to statement 1 above, a solution can only be unique if the echelon form has a pivot in every column. This is impossible if $m > n$. ∎

Theorem 2.2.2. *A matrix A is invertible if and only if its echelon form has pivot in every column and every row.*

Proof. Since a matrix must have unique solution for $Ax = b$ for any b in order to be invertible, it is necessary that the echelon form has pivot in every column and row, according to statements 1 and 2 above. ∎

This directly implies that an invertible matrix **must be square**.
Since an invertible matrix must be square and must have pivots in every row and column in echelon form, any invertible matrix is row equivalent to the identity matrix.
We can use this to get the following algorithm for finding A^{-1}:

1. Form an *augmented* $n \times 2n$ matrix $\left[A \;\middle|\; I\right]$.
2. Perform row operations to transform A into I.
3. The matrix that was originally I will now be A^{-1}.

To fully understand this algorithm, remember that every row operation can be expressed as the left multiplication by an elementary matrix. Let $E = E_n \cdots E_2 E_1$ represent all the performed row operations. Since we know E transforms A to the identity matrix, we have $EA = I$, so $E = A^{-1}$. Since row operations affect the entire augmented matrix, we have $\left[A \;\middle|\; I\right] \to \left[EA \;\middle|\; EI\right] = \left[I \;\middle|\; A^{-1}\right]$.

2.3 Fundamental Subspaces

Definition 2.3.1. A **subspace** of vector space V is a non-empty subset $V_0 \subset V$ which is also a vector space. Subspaces must be non-empty because all vector spaces must contain the zero vector.

For any linear transformation $A : V \to W$, we can associate the following subspaces:

1. The *null space*, or *kernel*, of A which consists of all vectors $v \in V$ such that $Av = 0$.
2. The *range* of A which is the set of all vectors $w \in W$ which can be represented as $w = Av$ for $v \in V$.

By the *column by coordinate rule*, we know that any vector in $Range(A)$ can be represented as a weighted sum of the column vectors of A, which is why the term Column Space is sometimes used to refer to Range.
In addition, we can consider the corresponding subspaces of the transposed matrix. The term *row space* is used to denote $Range(A^T)$, and the term *left null space* is used to denote $Null(A^T)$. Together, these four subspaces are known as the **fundamental subspaces** of the matrix A.

Definition 2.3.2. The **dimension** of a vector space V, denoted $dim(V)$, is the number of vectors in a basis.

Theorem 2.3.1 (General solution of a linear equation)**.** *Let a vector x_1 denote a solution to the equation $Ax = b$, and let H be the set of all solutions of $Ax = 0$. Then the set*

$$x = x_1 + x_h : x_h \in H$$

is the set of all solutions of the equation $Ax = b$.
In other words,

$$\Big(\textit{General solution of } Ax = b\Big) = \Big(\textit{A particular solution of } Ax = b\Big) + \Big(\textit{General solution of } Ax = 0\Big)$$

Proof. We know $Ax_1 = b$ and $Ax_h = 0$. For $x = x_1 + x_h$,

$$Ax = A(x_1 + x_h) = Ax_1 + Ax_h = b + 0 = b$$

Therefore, any solution x for $Ax = b$ can be represented as $x = x_1 + x_h$ with some $x_h \in H$. ∎

The power of this theorem is its generality – it applies to all linear equations. Aside from showing the structure of the solution set, this theorem allows us to separate investigations of uniqueness from existence. To study uniqueness of a solution, we only need to analyze uniqueness of $Ax = 0$, which always has a solution.

Theorem 2.3.2. *In order to compute the fundamental subspaces, we need to do row reduction. Let A be the original matrix and let A_e be its echelon form.*

1. *The pivot columns of the original matrix A (ie the columns where after row operations we will have pivots in echelon form) give us a basis for $Range(A)$.*
2. *The pivot rows of A_e give us the basis in row space.*
3. *To find $Null(A)$, we need to solve $Ax = 0$.*

Proof. In turn,

1. We know the pivot columns of A_e form a basis for $Range(A_e)$. Since $A_e = EA$ (E is the matrix product of the elementary matrices representing the row operations completed), $A = E^{-1}A_e$. This means the corresponding columns in A of A_e is a basis of A.
2. We know that the pivot rows of the echelon form are linearly independent. Now we need only prove that they span the entirety of the row space. Notice that *row operations do not change the row space*. To prove this,
$$A_e = EA$$
where A is $m \times n$ and E is an $m \times m$ invertible matrix.
$$Range(A_e^T) = Range(A^T E^T) = A^T(Range(E^T)) = A^T(\mathbb{R}^m) = Range(A^T)$$
where the final step follows from applying an $n \times m$ matrix to $\mathbb{R}^m$, which is just a transformation from $\mathbb{R}^m$ to $Range(A^T)$.
3. Solving for $Ax = 0$ certainly gives us a spanning set for $Null(A)$. To prove the set is linearly independent, multiply each vector by its corresponding free variable and add. For every free variable x_k, the entry k is exactly x_k, so the only way the sum of the set is 0 is if all the free variables are 0.

■

As an example of these computations, consider the matrix

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

Performing row operations, we get the echelon form

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

So the first and third columns of the *original matrix* give us a basis for $Range(A)$:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \\ -1 \end{bmatrix}$$

We also know the basis for $Row(A)$ is the first and second row of the *echelon form*:

$$\begin{bmatrix}1\\1\\2\\2\\1\end{bmatrix}, \begin{bmatrix}0\\0\\-3\\-3\\-1\end{bmatrix}$$

To find $Null(A)$ we solve $Ax = 0$. The reduced echelon form is

$$\begin{bmatrix}1 & 1 & 0 & 0 & \frac{1}{3}\\0 & 0 & 1 & 1 & \frac{1}{3}\\0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\end{bmatrix}$$

This means

$$\begin{Bmatrix}x_1 = -x_2 - \frac{1}{3}x_5\\ x_2 \text{ is free}\\ x_3 = -x_4 - \frac{1}{3}x_5\\ x_4 \text{ is free}\\ x_5 \text{ is free}\end{Bmatrix} \longrightarrow x_2\begin{bmatrix}-1\\1\\0\\0\\0\end{bmatrix} + x_4\begin{bmatrix}0\\0\\-1\\1\\0\end{bmatrix} + x_5\begin{bmatrix}-\frac{1}{3}\\0\\-\frac{1}{3}\\0\\1\end{bmatrix}$$

The vectors at each free variables form the basis for $Null(A)$.

Definition 2.3.3. The **rank** of a linear transformation A, denoted $rank(A)$, is the dimension of the range of A.

$$rank(A) := Dim(Range(A))$$

Theorem 2.3.3 (The Rank Theorem)**.** *For a matrix A*

$$rank(A) = rank(A^T)$$

The proof of this is trivial since rank of both column space and row space are dependent on the number of pivots in echelon form.

Theorem 2.3.4. *Let A be an $m \times n$ matrix. Then*

1. $dim(Null(A)) + dim(Range(A)) = n$ *(dim of domain)*
2. $dim(Null(A^T)) + dim(Range(A^T)) = dim(Null(A^T)) + rank(A) = m$ *(dim of codomain)*

Proof. In turn,

1. The first equality is simply that the number of free variables + the number of pivots = the number of columns.
2. The second equality applies the Rank Theorem to prove the row counterpart to the first equality.

■

The following follows from the second statement in the above theorem.

Theorem 2.3.5. *Let A be an $m \times n$ matrix. Then the equation*

$$Ax = b$$

has a solution for every $b \in \mathbb{R}^m$ if and only if the dual equation

$$A^T x = 0$$

has only the trivial solution.

2.4 Change of Basis

Let V be a vector space with a basis $B := b_1, \cdots, b_n$. Recall that any vector $v \in V$ can be written

$$v = x_1 b_1 + \cdots + x_n b_n$$

where the numbers $x_1, \cdots, x_n$ are called the coordinates of v. We can write the *coordinate vector* as

$$[v]_B := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n$$

Note that $v \mapsto [v]_B$ is an isomorphism between V and $\mathbb{F}^n$.

Definition 2.4.1. Let $T : V \to W$ be a linear transformation, and let $A = \{a_1, \cdots, a_n\}$, $B = \{b_1, \cdots, b_m\}$ be bases in V, W respectively.
A **matrix of transformation T in bases A and B** is an $m \times n$ matrix, denoted by $[T]_{BA}$,

$$[Tv]_B = [T]_{BA}[v]_A$$

The matrix $[T]_{BA}$ is easy - its kth column is just $[Ta_k]_B$.

Definition 2.4.2. For the above two bases A and B, the **change of basis** is

$$[v]_B = [I]_{BA} v_A$$

where $[I]_{BA}$ is the **change of basis matrix** whose kth column is $[a_k]_B$.
Clearly, any change of basis is invertible and

$$[I]_{BA} = ([I]_{AB})^{-1}$$

Definition 2.4.3. We can use this to define **similar matrices** as matrices A, B such that

$$A = Q^{-1} B Q$$

This means we can treat similar matrices as different representations of the same linear operator.

Chapter 3

Determinants

3.1 Properties

In this chapter, we only consider determinants of $n \times n$ matrices. We will think of the determinant as the n-dimensional volume of the parallelepiped determined by our n vectors, $v_1, \cdots, v_n$. For dimensions 2 and 3, "volume" of the parallelepiped is determined with the *base times height* rule: we pick one vector and define height to be the distance between this vector and the subspace spanned by the $n-1$ remaining vectors. Then we define base to be the $(n-1)$-dimensional volume of the parallelepiped determined by the $n-1$ vectors. This understanding allows determinants the following properties:

1. **Linearity in each argument:** Multiplying some vector v_k by α means the height is multiplied by α which means the determinant is multiplied by the same constant. This means the determinant is *linear in each argument*, which means if we fix $n-1$ vectors the determinant is linear with respect to the final vector.

   Linearity means that for an $n \times n$ matrix A, $det(\alpha A) = \alpha^n det(A)$, because multiplying A by α is equivalent to multiplying n columns by α.

2. **Preservation under column replacement:**

   $$det(v_1, \cdots, v_j + \alpha v_k, \cdots, v_k, \cdots, v_n) = det(v_1, \cdots, v_j, \cdots, v_k, \cdots, v_n)$$

   This is true because the "height" of $v_j + \alpha v_k$ is the same as the "height" of v_j, since "height" is defined in relation to the distance from the remaining subspace.

3. **Antisymmetry:** Swapping two vectors means the determinant changes signs.

   $$det(v_1, \cdots, v_k, \cdots, v_j, \cdots, v_n) = -det(v_1, \cdots, v_j, \cdots, v_k, \cdots, v_n)$$

   This does not seem natural at first, but we can prove it by applying preservation under column replacement thrice and then linearity.

   $$\begin{aligned}
   &det(v_1, \cdots, v_k, \cdots, v_j, \cdots, v_n) \\
   &= det(v_1, \cdots, v_k, \cdots, v_j - v_k, \cdots, v_n) \\
   &= det(v_1, \cdots, v_k + (v_j - v_k), \cdots, v_j - v_k, \cdots, v_n) \\
   &= det(v_1, \cdots, v_j, \cdots, v_j - v_k - (v_j), \cdots, v_n) \\
   &= det(v_1, \cdots, v_j, \cdots, -v_k, \cdots, v_n) \\
   &= -det(v_1, \cdots, v_j, \cdots, v_k, \cdots, v_n)
   \end{aligned}$$

4. **Normalization:** For the standard basis, the corresponding parallelepiped is the n-dimensional unit cube so its volume is 1.

$$det(I) = 1$$

Using these, we can derive additional basic properties of determinants for a square matrix A:

1. If A has a zero column, then $det(A) = 0$.
2. If A has two equal columns, then $det(A) = 0$.
3. If one column of A is a multiple of another, then $det(A) = 0$.

3.2 Computing the Determinant

The **determinant of diagonal matrices** is the product of the diagonal entries. Note that any diagonal matrix $\{a_1, \cdots, a_k\}$ can be obtained by multiplying column k of the identity matrix by a_k.
The **determinant of triangular matrices** is also the product of the diagonal entries. This is because an upper or lower triangular matrix can be reduced to a diagonal matrix with the same diagonal entries through column operations.

Theorem 3.2.1. *$det(A) = 0$ if and only if A is not invertible.*

Proof. Recall that we can only use column operations when reducing a matrix to find the determinant, which is equivalent to doing row operations on A^T. If the echelon form of A^T does not have pivots in every column and row, then the product of diagonal entries will be 0. Not having pivots in every column and row also means the matrix is not invertible, so the two conclusions are equivalent. ■

We will now prove some nontrivial properties of determinants, but to do so we will need the following two lemmas.

Lemma 3.2.2. *For a square matrix A and elementary matrix E,*

$$det(AE) = det(A)det(E)$$

Proof. Right multiplication of an elementary matrix is simply a column operation. Since a column operation is obtained from the identity matrix by the column operation, its determinant is 1 times the effect of the column operation. ■

Lemma 3.2.3. *Any invertible matrix is a product of elementary matrices.*

Proof. We know that any invertible matrix is *row equivalent* to the identity matrix, which is its reduced echelon form. So

$$I = E_n E_{n-1} \cdots E_1 A$$

which means we can write A in terms of the identity and the inverses of some elementary matrices

$$A = E_1^{-1} \cdots E_{n-1}^{-1} E_n^{-1} I = E_1^{-1} \cdots E_{n-1}^{-1} E_n^{-1}$$

Since the inverse of an elementary matrix is an elementary matrix, the proof is complete. ■

Now for two important theorems:

Theorem 3.2.4. *For a square matrix A,*

$$det(A) = det(A^T)$$

Proof. A key observation is that $det(E) = det(E^T)$ for any elementary matrix E.
Notice also that it is sufficient to prove the theorem *only* for *invertible matrices* since if A is not invertible then A^T is also not invertible and both determinants are 0, trivially proving the theorem.
Now, by the above lemma we can write

$$A = E_1 E_2 \cdots E_n$$

which means

$$det(A) = det(E_1)det(E_2)\cdots det(E_n)$$

We can also write

$$A^T = E_n^T \cdots E_2^T E_1^T = E_n \cdots E_2 E_1$$

which means

$$det(A^T) = det(E_n)\cdots det(E_2)det(E_1)$$

which is equivalent to $det(A)$. ■

This theorem means that column operations have the same effect on determinants as row operations, so we can use either when reducing matrices to compute determinants.

Theorem 3.2.5. *For $n \times n$ matrices A, B,*

$$det(AB) = det(A)det(B)$$

Proof. Two cases:
Case 1: B is invertible.
This means we can write

$$B = E_1 E_2 \cdots E_n$$

and so

$$det(AB) = det(A)[det(E_1)det(E_2)\cdots det(E_n)] = det(A)det(B)$$

Case 2: B is not invertible. If B is not invertible, we will prove that the product AB is also not invertible so $det(AB) = det(A)det(B)$ simplifies to $0 = 0$.
We proceed by contradiction. Assume $AB = C$ is invertible. Then we left multiply both sides by C^{-1} to get $C^{-1}AB = I$, which means $C^{-1}A$ is the left inverse of B, but because B is square, that means $C^{-1}A$ is the inverse of B. Since we know B is not invertible, we have a contradiction. ■

3.3 Cofactor Expansion

For an $n \times n$ matrix A, let $A_{j,k}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by crossing out row j and column k.

Theorem 3.3.1 (Cofactor expansion of determinant)**.** *For each $j, 1 \leq j \leq n$, the determinant of A can be expanded in the row number j as*

$$det(A) = a_{j,1}(-1)^{j+1} det(A_{j,1}) + a_{j,2}(-1)^{j+2} det(A_{j,2}) + \cdots + a_{j,n}(-1)^{j+n} det(A_{j,n})$$

A similar expansion can be done for columns.

Proof. We will prove the expansion for row 1. This can be generalized by swapping row 1 with another row. Additionally, since $det(A) = det(A^T)$, column expansion follows automatically.
Consider the special case when the first row has *only one* nonzero term, $a_{1,1}$. Performing column operations on columns $2, \cdots, n$, we transform A to lower triangular form. Now

$$det(A) = (\text{product of diagonal entries}) \times (\text{correcting factor from column operations})$$

but since the *product of diagonal entries* except $a_{1,1}$ times the *correcting factor* is exactly $det(A_{1,1})$, we can write

$$det(A) = a_{1,1} det(A_{1,1})$$

Now consider the case when all entries in the first row except $a_{1,2}$ are zeros. We can reduce this to the previous case by swapping columns 1 and 2, so $det(A) = (-1)a_{1,2}det(A_{1,2})$.
If $a_{1,3}$ is the only nonzero term in the first row, we can reduce this to the previous case by swapping columns 2 and 3, so $det(A) = a_{1,3}det(A_{1,3})$. We do this instead of swapping columns 1 and 3 to maintain the order of the $n-1$ other columns.
These special cases are important because we have linearity of the determinant. If the matrix $A^{(k)}$ is obtained by replacing all A's entries in the first row with 0 except for $a_{1,k}$, then linearity of the determinant implies

$$det(A) = det(A^{(1)}) + \cdots + det(A^{(n)}) = \sum_{k=1}^{n} det(A^{(k)})$$

Based on our analysis of special cases, we know

$$det(A^{(k)}) = (-1)^{1+k} a_{1,k} det(A_{1,k})$$

so

$$det(A) = \sum_{k=1}^{n} (-1)^{1+k} a_{1,k} det(A_{1,k})$$

To get the expansion for the second row, we swap rows so multiply by -1. For the third row, multiply by -1 again to get the original equation, and so on. ∎

Cofactor expansion is not practical for anything larger than a 3×3 matrix, but it has great theoretical importance.

Definition 3.3.1. Formally, the numbers

$$C_{j,k} = (-1)^{j+k} det(A_{j,k})$$

are called **cofactors**.
The matrix $C = \{C_{j,k}\}_{j,k=1}^{n}$ whose entries are *cofactors* of a given matrix A is called the **cofactor matrix** of A.

Theorem 3.3.2 (Cofactor formula for inverse)**.** *Let A be an invertible matrix and let C be its cofactor matrix. Then*

$$A^{-1} = \frac{1}{det(A)} C^T$$

Proof. Let us find the product AC^T.
The jth diagonal entry is obtained by multiplying the jth row of A by the jth row of C,

$$(AC^T)_{j,j} = a_{j,1}C_{j,1} + \cdots + a_{j,n}C_{j,n} = det(A)$$

by cofactor expansion.
To get the off-diagonal terms, we multiply the kth row of A with the jth row of C, $j \neq k$,

$$a_{k,1}C_{j,1} + \cdots + a_{k,n}C_{j,n}$$

If we look at this as a cofactor expansion of the jth row, this is the determinant of the matrix A except that we replace row j with row k. Since two rows of our matrix coincide, the determinant will be 0, which means all off-diagonal terms will be 0, thus

$$AC^T = det(A)I$$

■

Since for invertible matrices, $Ax = b$ has a unique solution, we have

$$x = A^{-1}b = \frac{C^T b}{det(A)}$$

Theorem 3.3.3 (Cramer's Rule)**.** *For invertible matrix A, entry k of the solution to $Ax = b$ is given by*

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

where B_k is obtained from A by replacing column k with b.

Proof. After our above theorem, we need only prove that entry k of $C^T b = det(B_k)$.
We know entry k of $C^T b$ is equivalent to the product of the kth row of C^T and b, which is equivalent to the product of the kth column of C and b.
C_{jk} is obtained by crossing out the jth row and kth column of A and computing the determinant of the remaining matrix. Multiplying the kth column of C with b is equivalent to

$$b_1 C_{1,k} + \cdots + b_n C_{n,k}$$

which is the same as the cofactor expansion of B_k. ■

One application of the cofactor formula is a shortcut to inverting 2×2 matrices. For the matrix

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

The cofactor matrix is made up of 4 individual 1×1 matrices,

$$C = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

which means

$$A^{-1} = \frac{1}{det(A)} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

Chapter 4

Spectral Theory

Spectral theory will be our main tool for analyzing linear operators. In this chapter, we only consider transformations $A : V \to V$ ($n \times n$ matrices).

4.1 Definitions

Definition 4.1.1. A scalar λ is called an **eigenvalue** of operator $A : V \to V$ if there exists a *nonzero* vector $v \in V$ such that

$$Av = \lambda v$$

The vector v is called an **eigenvector** of A (corresponding to the eigenvalue of λ).
Once we know the eigenvalues, finding the eigenvectors is equivalent to solving

$$(A - \lambda I)v = 0$$

$Null(A - \lambda I)$, the set of all eigenvectors and 0, is called the **eigenspace**.
The set of all eigenvalues of an operator is called the **spectrum** of A, denoted $\sigma(A)$.

Since the matrix A is square, $A - \lambda I$ has a nontrivial null space if and only if it is not invertible, which means its determinant will be 0. Thus, for any eigenvalue λ of A,

$$det(A - \lambda I) = 0$$

Definition 4.1.2. If A is an $n \times n$ matrix, $det(A - \lambda I)$ is a degree-n polynomial of variable λ. This is called the **characteristic polynomial** of A. Finding the spectrum of A requires finding the roots to the characteristic polynomial.
Using $(\lambda I - A)v = 0$ as the characteristic polynomial always yields a monic polynomial, whereas our current definition differs by a factor of $(-1)^n$. This makes no difference for properties like having eigenvalues located at roots so the two definitions are usually interchangeable.
This means any operator in $\mathbb{C}^n$ has n eigenvalues, though some may be repeated.

Theorem 4.1.1. *An $n \times n$ matrix A is invertible if and only if it doesn't have an eigenvalue of 0.*

Proof. Proving if:
If A doesn't have an eigenvalue of 0, then $det(A-0I) \neq 0 \rightarrow det(A) \neq 0$, which implies A is invertible.
Proving only if:
If A is invertible, then $det(A) \neq 0$, which implies $det(A-0I) \neq 0$. ■

Theorem 4.1.2. *Let A be an $n \times n$ matrix, and let $\lambda_1, \cdots, \lambda_n$ be its complex eigenvalues (counting multiplicities). Then*

$$det(A) = \lambda_1 \cdots \lambda_n$$

Proof. Since $det(A-\lambda I)$ is a degree-n polynomial of variable λ and we know A will have n eigenvalues, we can write

$$det(\lambda I - A) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

Plugging in $\lambda = 0$ gives us

$$det(-A) = (-1)^n det(A) = (-\lambda_1) \cdots (-\lambda_n)$$

which simplifies to

$$det(A) = (\lambda_1) \cdots (\lambda_n)$$

■

Theorem 4.1.3. *Let A be an $n \times n$ matrix, and let $\lambda_1, \cdots, \lambda_n$ be its complex eigenvalues (counting multiplicities). Then*

$$trace(A) = \lambda_1 + \cdots + \lambda_n$$

Proof. Let us begin by analyzing $det(A - \lambda I)$. Notice that in any cofactor expansion, if we pick any element $a_{i,j}$, such that $j \neq k$, then the highest degree of the resulting cofactor will be $n-2$. This is because cofactoring removes the row and column the chosen entry is on, and since $j \neq k$, we remove the variables $a_{j,j} - \lambda$ and $a_{k,k} - \lambda$. After cofactor expansion, the λ^{n-1} term will be formed by only this equation

$$(a_{1,1} - \lambda) \cdots (a_{n,n} - \lambda) = (-1)^n (\lambda - a_{1,1}) \cdots (\lambda - a_{n,n})$$

so the coefficient of λ^{n-1} amounts to choosing the λ variable $n-1$ times and choosing one of the other coefficients to get

$$(-1)^n (a_{1,1}\lambda^{n-1}) \cdots (a_{n,n}\lambda^{n-1}) = (-1)^n (a_{1,1} + \cdots + a_{n,n})\lambda^{n-1} \tag{4.1}$$

Note we can rewrite the characteristic equation as

$$det(A - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

Now let us identify the coefficient of the λ^{n-1} term

$$(-1)^n (\lambda_1 \lambda^{n-1} + \lambda_2 \lambda^{n-1} + \cdots + \lambda_n \lambda^{n-1}) = (-1)^n (\lambda_1 + \cdots + \lambda_n)\lambda^{n-1} \tag{4.2}$$

Comparing coefficients in Equations 4.1 and 4.2, $trace(A) = \lambda_1 + \cdots + \lambda_n$. ■

4.2 Diagonalization

We can use spectral theory to find the diagonalization of operators, which means that given an operator, we find the basis in which the matrix of the operator is diagonal. This makes powers of an operator much easier

to compute.

Theorem 4.2.1. *A matrix A in $\mathbb{F}^n$ can be written as $A = PDP^{-1}$, where D is a diagonal matrix and P is invertible, if and only if there exists a basis in $\mathbb{F}^n$ of eigenvectors of A.*
In this case, the diagonal entries of D are the eigenvalues of A and the columns of P are the corresponding eigenvectors.

Proof.
To understand the intuition behind this, note that $P = [I]_{S,B}$, where S is the standard basis and B is the basis for the eigenspace, since each column is the representation of a basis vector written in S. Rewriting $A = PDP^{-1}$ as $D = P^{-1}AP = [I]_{B,S}A[I]_{S,B}$ which means $D = [A]_{B,B}$, which is a diagonal operator if and only if its diagonal entries are eigenvalues whose corresponding eigenvectors are b_k. Think of the operator $[I]_{B,S}A[I]_{S,B}$ as converting a vector to a basis of eigenvectors, scaling those eigenvectors appropriately by their eigenvalues, and then converting back to the standard basis.
A simpler, more direct proof is to rewrite $AP = PD$.

$$AP = \begin{bmatrix} Ab_1 & \cdots & Ab_n \end{bmatrix} = \begin{bmatrix} \lambda_1 b_1 & \cdots & \lambda_n b_n \end{bmatrix}$$

$$PD = \begin{bmatrix} b_1\lambda_1 & \cdots & b_1\lambda_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 b_1 & \cdots & \lambda_n b_n \end{bmatrix}$$

■

Of course, for P to be invertible, the eigenvectors $b_1, \cdots, b_n$ must be linearly independent. Luckily, we can easily check if this is the case with the following theorem.

Theorem 4.2.2. *Let $\lambda_1, \cdots, \lambda_n$ be **distinct** eigenvalues for A, and let $b_1, \cdots, b_n$ be their corresponding eigenvectors. Then $b_1, \cdots, b_n$ are linearly independent.*

Proof. We proceed by induction over the n eigenvectors of A.
Base case: $n = 1$
This is trivial because by definition, an eigenvector is nonzero. Any set consisting of a single nonzero vector is linearly independent.
Inductive Hypothesis:
Assume it holds for $n = k$.
Inductive Step: $n = k + 1$
Suppose there exists a non-trivial solution to

$$\sum_{i=1}^{k+1} c_i b_i = 0$$

We can apply $(A - \lambda_{k+1}I)$ to both sides to get

$$\sum_{i=1}^{k+1} c_i (A - \lambda_{k+1}I) b_i = 0$$

Since $(A - \lambda_{k+1}I)b_{k+1} = 0$, we can write

$$\sum_{i=1}^{k} c_i (A - \lambda_{k+1}I) b_i = \sum_{i=1}^{k} c_i (\lambda_i - \lambda_{k+1}) b_i = 0$$

By the inductive hypothesis, we know the first k eigenvectors are linearly independent, so the coefficient $c_i(\lambda_i - \lambda_{k+1})$ must be 0 for $0 \leq i \leq k$, and since eigenvalues are distinct, $c_i = 0$ for $0 \leq i \leq k$.
Now we can reduce our original summation

$$\sum_{i=1}^{k+1} c_i b_i = c_{k+1} b_{k+1} = 0$$

This means that c_{k+1} must be 0, so the summation only has the trivial solution, which means the eigenvectos are linearly independent. ■

Chapter 5

Inner Product Spaces

Keep in mind that theory for inner product space is only developed for $\mathbb{R}$ and $\mathbb{C}$, so $\mathbb{F}$ will always denote one of those two fields in the next two chapters.

5.1 Inner Product

Definition 5.1.1. We define the **norm** of a vector to be the generalization of *length*. That is, the norm of a vector $x \in \mathbb{R}^n$ is

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

For any complex number $z = x + iy$, we can write $|z|^2 = x^2 + y^2 = z\bar{z}$, where $\bar{z}$ denotes the complex conjugate of z. So for any z in a complex field $\mathbb{C}^n$, we can write

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 + iy_1 \\ \vdots \\ x_n + iy_n \end{bmatrix}$$

so it is natural to define the norm $\|z\|$ as

$$\|z\|^2 = \sum_{k=1}^{n} (x_k^2 + y_k^2) = \sum_{k=1}^{n} |z_k|^2$$

Definition 5.1.2. The **inner product** of two vectors $x, y \in \mathbb{R}^n$ is

$$(x, y) = x_1 y_1 + \cdots + x_n y_n = x^T y = y^T x$$

This yields another definition for the **norm**:

$$\|x\| = \sqrt{(x, x)}$$

For complex fields, we need a definition of inner product such that $\|z\|^2 = (z, z)$. One definition that is consistent with this requirement will be our definition for the ***standard*** **inner product in** $\mathbb{C}^n$,

$$(z, w) = z_1 \overline{w_1} + \cdots + z_n \overline{w_n}$$

To simplify this, we will define the **Hermitian adjoint**, or simply **adjoint** A^*, by $A^* = \overline{A}^T$.

Using this, we can write

$$(z, w) = w^* z$$

The inner products we defined for $\mathbb{R}^n$ and $\mathbb{C}^n$ have the following properties:

1. Symmetry: $(x, y) = \overline{(y, x)}$
2. Linearity: $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$
3. Non-negativity: $(x, x) \geq 0$
4. Non-degeneracy: $(x, x) = 0$ if and only if $x = 0$

Note that properties 1 and 2 imply that

$$(x, \alpha y + \beta z) = \overline{(\alpha y + \beta z, x)} = \overline{\alpha}(x, y) + \overline{\beta}(x, z)$$

Lemma 5.1.1. *Let x be a vector in V. Then $x = 0$ if and only if*

$$(x, y) = 0 \qquad \forall y \in V$$

Proof. Since $(0, y) = 0$, we need to only show that $x = 0$ if $(x, y) = 0$. Subbing in $y = x$, we get $(x, x) = 0$ and property 3 asserts that $x = 0$. ■

Lemma 5.1.2. *Let x, y be vectors in V. Then $x = y$ if and only if*

$$(x, z) = (y, z) \qquad \forall z \in V$$

Proof. Using the above lemma, if we set $(x - y, z) = 0 \quad \forall z \in V$, then it follows that $x = y$ and $(x, z) = (y, z)$. ■

Theorem 5.1.3. *Suppose two operators $X, Y : A \to B$ satisfy*

$$(Ax, y) = (Bx, y) \qquad \forall x \in X, \forall y \in Y$$

Then $A = B$.

Proof. Using the previous lemma, we can fix x and take all $y \in Y$, which means $Ax = Bx$. Since this is true for all x, A and B are the same operator. ■

Theorem 5.1.4 (Cauchy-Schwarz Inequality)**.**

$$\left|(x, y)\right| \leq \|x\| \cdot \left\|y\right\|$$

Proof. If x or y is 0, then the proof is trivial. Assuming neither is 0, we will prove both the real and complex cases. But first consider only the real case:

$$0 \leq \left\|x - ty\right\|^2 = (x - ty, x - ty) = \|x\|^2 - 2t(x, y) + t^2\left\|y\right\|^2$$

Taking the derivative with respect to t and setting it to 0 gives us $t = \frac{(x,y)}{\|y\|^2}$. We will use this same t value for the following proof of the real and complex cases:

$$\begin{aligned} 0 \le \left\|x - ty\right\|^2 &= (x - ty, x - ty) \\ &= (x, x - ty) - t(y, x - ty) \\ &= \|x\|^2 - \bar{t}(x, y) - t(y, x) + |t|^2 \left\|y\right\|^2 \end{aligned}$$

Using property 1 of inner products, we have

$$t = \frac{(x, y)}{\left\|y\right\|^2} = \frac{\overline{(y, x)}}{\left\|y\right\|^2}$$

Subbing in t, we get

$$0 \le \|x\|^2 - \frac{\left|(xy)\right|^2}{\left\|y\right\|^2}$$

which completes the proof. ■

Theorem 5.1.5 (Triangle Inequality)**.**

$$\left\|x + y\right\| \le \|x\| + \left\|y\right\|$$

Proof.

$$\begin{aligned} \left\|x + y\right\|^2 = (x + y, x + y) = \|x\|^2 + \left\|y\right\|^2 + (x, y) + (y, x) \\ &\le \|x\|^2 + \left\|y\right\|^2 + 2\left|(x, y)\right| \\ &\le \|x\|^2 + \left\|y\right\|^2 + 2\|x\| \cdot \left\|y\right\| \\ &= (\|x\| + \left\|y\right\|)^2 \end{aligned}$$

■

Theorem 5.1.6. *The following **polarization identities** allow us to construct the inner product from the norm:*
For $x, y \in \mathbb{R}^n$,

$$(x, y) = \frac{1}{4}\left(\left\|x + y\right\|^2 - \left\|x - y\right\|^2\right)$$

For $x, y \in \mathbb{C}^n$,

$$(x, y) = \frac{1}{4}\left(\left\|x + y\right\|^2 - \left\|x - y\right\|^2 + i\left\|x + iy\right\|^2 - i\left\|x - iy\right\|^2\right)$$

Proof. For the real case,

$$\begin{aligned} \left\|x + y\right\|^2 - \left\|x - y\right\|^2 &= (x + y, x + y) - (x - y, x - y) \\ &= \|x\|^2 + \left\|y\right\|^2 + 2(x, y) - \|x\|^2 - \left\|y\right\|^2 + 2(x, y) \\ &= 4(x, y) \end{aligned}$$

For the complex case,

$$\begin{aligned}\sum_{k=0}^{3} i^k \left\|x + i^k y\right\|^2 &= \sum_{k=0}^{3} i^k (x + i^k y, x + i^k y) \\ &= \sum_{k=0}^{3} i^k \left(\|x\|^2 + \left\|y\right\|^2 + (x, i^k y) + (i^k y, x) \right) \\ &= \sum_{k=0}^{3} \left(i^k \|x\|^2 + i^k \left\|y\right\|^2 + (x, y) + (i^{2k} y, x) \right) \\ &= 4(x, y)\end{aligned}$$

where the last step follows from

$$\sum_{k=0}^{3} i^k = \sum_{k=0}^{3} i^{2k} = 0$$

■

Theorem 5.1.7 (Parallelogram Identity)**.** *Another important property of the norm is the parallelogram identity. For vectors u and v:*

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$$

Proof. The theorem follows easily from the fact that the sum of the diagonals of a parallelogram equal the sum of all four sides. ■

To review, we have just proved the following properties about the norm $\|u\|$:

1. Homogeneity: $\|\alpha u\| = |\alpha| \cdot \|u\|$
2. Triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$
3. Non-negativity: $\|u\| \geq 0$
4. Non-degeneracy: $\|u\| = 0$ if and only if $u = 0$

In a vector space V, if we assign to each vector u a number $\|u\|$ that satisfies these 4 properties, we can say that the space V is a **normed space**.

5.2 Orthogonality

Definition 5.2.1. Two vectors u and v are **orthogonal**, denoted $u \perp v$, if and only if $(u, v) = 0$

Theorem 5.2.1. *If $u \perp v$, then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Proof.

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2 + (u, v) + (v, u) = \|u\|^2 + \|v\|^2$$

Since $(u, v) = (v, u) = 0$ because of orthogonality. ■

Definition 5.2.2. A vector u is **orthogonal to vector space** V if u is orthogonal to all vectors in V.

Theorem 5.2.2. *Let V be spanned by $v_1, \cdots, v_n$. Then $u \perp V$ if and only if*

$$u \perp v_k \qquad \forall k = 1, \cdots, n$$

Proof. Proving "only if" is trivial by the definition of $u \perp V$. Proving "if" comes easily after noticing that any vector can be rewritten as a linear combination of the basis vectors, so if u is perpendicular to all the basis vectors, then it is perpendicular to any other vector in V. ■

Definition 5.2.3. A set of vectors $v_1, \cdots, v_n$ are orthogonal if any two vectors in the set are orthogonal to each other. If $\|v_k\| = 1$ for all k, we call the set orthonormal.

Lemma 5.2.3 (Generalized Pythagorean Theorem)**.** *Let $v_1, \cdots, v_n$ be an orthogonal system. Then*

$$\left\| \sum_{k=1}^{n} a_k v_k \right\|^2 = \sum_{k=1}^{n} |a_k|^2 \|v_k\|^2$$

Proof.

$$\left\| \sum_{k=1}^{n} a_k v_k \right\|^2 = \left(\sum_{k=1}^{n} a_k v_k, \sum_{j=1}^{n} a_j v_j \right) = \sum_{k=1}^{n} \sum_{j=1}^{n} a_k \overline{a_j} (v_k, v_j)$$

Since the set is orthogonal, (v_k, v_j) is only nonzero when $k = j$, so

$$= \sum_{k=1}^{n} |a_k|^2 \|v_k\|^2$$

■

Definition 5.2.4. An orthogonal set of vectors that is also a basis is called an **orthogonal basis**.

Typically, to find coordinates of a vector in a basis, we need to solve a system of equations. For orthogonal bases, it is much simpler. Suppose $v_1, \cdots, v_n$ is an orthogonal basis and let

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

Taking the inner product with v_1 yields

$$(x, v_1) = (\sum_{j=1}^{n} \alpha_j (v_j, v_1) = \alpha_1 (v_1, v_1) = \alpha_1 \|v_1\|^2$$

Thus, to find any coordinate α_k of a vector x in orthogonal basis $v_1, \cdots, v_n$:

$$\alpha_k = \frac{(x, v_k)}{\|v_k\|^2}$$

This is a simple example of abstract orthogonal Fourier decomposition – simple because classical Fourier decomposition deals with infinite orthonormal systems.

5.3 Orthogonal Projection and Gram-Schmidt Orthogonalization

Definition 5.3.1. The **orthogonal projection** of a vector v onto the subspace E is the vector $w := P_E v$ such that $w \in E$ and $v - w \perp E$.

Theorem 5.3.1. *The orthogonal projection $w = P_E v$ minimizes the distance from v to E. In other words,*

$$\|v - w\| \leq \|v - x\| \qquad \forall x \in E$$

Additionally, if for some $x \in E$

$$\|v - w\| = \|v - x\|$$

then $x = w$.

Proof. Let $y = w - x \in E$. Then

$$v - x = v - w + w - x = v - w + y$$

Since $v - w \perp E$, we know $v - w \perp y$. By the Pythagorean Theorem,

$$\|v - x\|^2 = \|v - w\|^2 + \|y\|^2 \geq \|v - w\|^2$$

To finish the proof, note that equality only arises when $y = 0$, ie when $x = w$. ∎

There is a formula for finding an orthogonal projection if we know an orthogonal basis in E. Let $v_1, \cdots, v_n$ be an orthogonal basis in E. Then the projection $P_E v$ of a vector v is

$$P_E v = \sum_{k=1}^{n} a_k v_k \qquad where \qquad a_k = \frac{(v, v_k)}{\|v_k\|^2}$$

In other words,

$$P_E v = \sum_{k=1}^{n} \frac{(v, v_k)}{\|v_k\|^2} v_k$$

Note that this only works if we have an *orthogonal* basis. However, even if we only have a basis in E, we can use the following algorithm to find an orthogonal basis.

Theorem 5.3.2 (Gram-Schmidt Orthogonalization Algorithm)**.** *Suppose we have linearly independent system $x_1, \cdots, x_n$. The Gram-Schmidt algorithm constructs from this an orthogonal system $v_1, \cdots, v_n$ such that*

$$span(x_1, \cdots, x_n) = span(v_1, \cdots, v_n)$$

Additionally, for all $r \leq n$

$$span(x_1, \cdots, x_r) = span(v_1, \cdots, v_r)$$

The algorithm is as follows:

1. *Define $v_1 := x_1$.*
   Define $E_1 := span(v_1) = span(x_1)$.
2. *Define $v_2 := x_2 - P_{E_1} x_2 = x_2 - \frac{(x_2, v_1)}{\|v_1\|^2} v_1$.*
   Define $E_2 := span(v_1, v_2) = span(x_1, x_2)$.
3. *Define $v_3 := x_3 - P_{E_2} x_3 = x_3 - \frac{(x_3, v_1)}{\|v_1\|^2} v_1 - \frac{(x_3, v_2)}{\|v_2\|^2} v_2$.*
   Define $E_3 := span(v_1, v_2, v_3) = span(x_1, x_2, x_3)$.

4. *Continue until we have n vectors and* $span(v_1,\cdots,v_n) = span(x_1,\cdots,x_n)$. *The formula for vector* v_{r+1} *given* $v_1,\cdots,v_r$ *is*

$$v_{r+1} := x_{r+1} - P_{E_r}x_{r+1} = x_{r+1} - \sum_{k=1}^{r} \frac{(x_{r+1}, v_k)}{\|v_k\|^2} v_k$$

Note that at each step, we are adding in x_{r+1} *which means the resulting vector will not exist in* E_r.

Proof. At each step, we add in x_{r+1} and then subtract its projection the subspace spanned by $x_1,\cdots,x_r$, meaning each additional vector is orthogonal to the ones previously defined. Since we set $v_1 = x_1$, we have proved the algorithm by induction. ■

Since multiplication by a scalar does not change orthogonality, we can multiply vectors v_k returned by Gram-Schmidt by any non-zero numbers. One use case is to normalize the orthogonal vectors by dividing by their norms $\|v_k\|$ to yield an orthonormal system.

Theorem 5.3.3. *A projector P is independent of the basis used to define the subspace W being projected onto.*

Proof. Suppose we are projecting $v \in V$ onto W. Define w_1 and w_2 as two projections of v onto W.
By definition, $(v - w_1) \perp W$ and $(v - w_2) \perp W$.

$$\begin{aligned}(v - w_2 - (v - w_1)) \perp W &\rightarrow (w_1 - w_2) \perp W \\ &\rightarrow (w_1 - w_2) \perp (w_1 - w_2) \\ &\rightarrow w_1 - w_2 = 0\end{aligned}$$

where the second step follows from $(w_1 - w_2) \in W$ and the third step follows from the fact that only the 0 vector is orthogonal to itself. This implies both projections are the same. Since this is true for all projections $P_W v$, P is unique. ■

The following is an important, defining property of projections.

Theorem 5.3.4. *For any orthogonal projection,* $P : V \rightarrow E$,

$$P^2 = P$$

Proof. We know that for some orthogonal basis $v_1,\cdots,v_n$ for E,

$$P_E v = \sum_{k=1}^{n} \frac{(v, v_k)}{\|v_k\|^2} v_k$$

If $v \in E$, then

$$v = \sum_{k=1}^{n} \frac{(v, v_k)}{\|v_k\|^2} v_k$$

so $P_E v = v$ for all $v \in E$.
Intuitively, projecting $v \in E$ onto E would clearly result in v. ■

The idea of orthogonality is not unique to vector spaces. As an example, we will use Gram-Schmidt to construct a set of *orthogonal polynomials* over the interval [-1,1]. Let us define the inner product on the space of polynomials as $(f, g) = \int_{-1}^{1} f(t)g(t)dt$. Now, we will use Gram-Schmidt to construct an orthogonal basis of polynomials from the basis $1, t, t^2, t^3$ such that for each polynomial $P(1) = 1$.
We know that

$$v_1 := 1$$

Then, we define

$$v_2 := t - \frac{(t,1)}{(1,1)}1 = t - \frac{\int_{-1}^{1} t\,dt}{\int_{-1}^{1} dt} = t$$

Next, we define

$$v_3 := t^2 - \frac{(t^2,1)}{(1,1)}1 - \frac{(t^2,t)}{(t,t}t = t^2 - \frac{\int_{-1}^{1} t^2 dt}{\int_{-1}^{1} dt} - \frac{\int_{-1}^{1} t^3 dt}{\int_{-1}^{1} t^2 dt} = \frac{1}{2}(3t^2 - 1)$$

Note that we scaled v_3 to satisfy $P(1) = 1$. Finally, we define

$$v_4 := t^3 - \frac{(t^3,1)}{(1,1)}1 - \frac{(t^3,t)}{(t,t)}t - \frac{(t^3,t^2)}{(t^2,t^2)}t^2 = t^3 - \frac{\int_{-1}^{1} t^4 dt}{\int_{-1}^{1} t^2 dt}t = \frac{1}{2}(5t^3 - 3t)$$

These are the first 4 of **Legendre's polynomials**, a set of important polynomials which appear in many different branches of mathematics.

5.3.1 Orthogonal Complement

Definition 5.3.2. For a subspace E, its **orthogonal complement** $E^\perp$ is the set of all vectors orthogonal to E. Since at least 0 is orthogonal to E, $E^\perp$ is always a subspace.

By the definition of orthogonal projection, any vector in an inner product space V has a unique representation of the form

$$v = v_1 + v_2 \qquad v_1 \in E, v_2 \in E^\perp$$

This statement is usually written as $V = E \oplus E^\perp$.

Theorem 5.3.5. *For subspace E of V,*

$$(E^\perp)^\perp = E$$

Proof. We will show $E \subseteq (E^\perp)^\perp$ and $(E^\perp)^\perp \subseteq E$.
Let $u \in E$. Then $(u,v) = 0$ for all $v \in E^\perp$. Since u is orthogonal to every vector $v \in E^\perp$, then $u \in (E^\perp)^\perp$ so $E \subseteq (E^\perp)^\perp$.
Now let $u \in (E^\perp)^\perp$. Since $V = E \oplus E^\perp$, we can write $u = v + w$, where $v \in E$ and $w \in E^\perp$. This means that $u - v = w \in E^\perp$. Since we know $E \subseteq (E^\perp)^\perp$, we have $u \in (E^\perp)^\perp$ and $v \in (E^\perp)^\perp$, which means $u - v \in (E^\perp)^\perp$. Therefore, $u - v \in E^\perp \cap (E^\perp)^\perp$. Since the only vector that is orthogonal to itself is 0, $u = v$, and because $v \in E$, $(E^\perp)^\perp \subseteq E$. ∎

Now suppose P is the orthogonal projection onto subspace E, and Q is the orthogonal projection onto the orthogonal complement $E^\perp$.

Lemma 5.3.6.

$$P + Q = I \text{ and } PQ = 0$$

Proof. For any inner product space V, we have $V = E \oplus E^\perp$. This means that for any vector $v \in V$, $P(v) = e_1$ where $e_1 \in E$, and so $(I - P)v = v - e_1 = e_2$ where $e_2 \in E^\perp$, which means $Q = I - P$.
Now,

$$P + Q = P + I - P = I$$

and

$$PQ = P(I - P) = P - P^2 = P - P = 0$$

■

Theorem 5.3.7. *$P - Q$ is its own inverse.*

Proof. We must show that $(P - Q)(P - Q) = I$.
Expanding

$$(P - Q)(P - Q) = P^2 - PQ - QP + Q^2 = P + Q = I$$

■

5.4 Least Square Solution

Recall that $Ax = b$ has a solution if and only if $b \in Range(A)$. In real life, it is impossible to avoid errors. The simplest way to approximate a solution is to choose an approximation $\hat{x}$ to minimize the error $e = \|A\hat{x} - b\|$. This is the **least square solution**.
We know $A\hat{x}$ is the orthogonal projection $P_{Range(A)}b$ if and only if $b - A\hat{x} \perp Range(A)$. Using the column space interpretation of range, this is equivalent to

$$b - A\hat{x} \perp a_k \qquad \forall k = 1, \cdots, n$$

That means

$$0 = (b - A\hat{x}, a_k) = a_k^*(b - A\hat{x}) \qquad \forall k = 1, \cdots, n$$

We can join the rows a_k^* together to get

$$A^*(b - A\hat{x}) = 0$$

which is equivalent to the **normal equation**

$$A^*A\hat{x} = A^*b$$

The solution $\hat{x}$ to this equation grants us the least square solution of $A\hat{x} = b$. This makes it easy to notice that the least square solution is unique if and only if A^*A is invertible.
If $\hat{x}$ is the solution to the normal equation, then $A\hat{x} = P_{Range(A)}b$. So in order to find the actual projection of b onto $Range(A)$, we need to solve the normal equation and then multiply the solution by A. Formally,

$$P_{Range(A)}b = A(A^*A)^{-1}A^*b$$

Because this is true for all b, the formula for the matrix of the orthogonal projection onto $Range(A)$ is

$$P_{Range(A)} = A(A^*A)^{-1}A^*$$

Theorem 5.4.1. *For an $m \times n$ matrix A*

$$Ker(A) = Ker(A^*A)$$

Recall Kernel is equivalent to Null Space.

Proof. We will show $Ker(A) \subseteq Ker(A^*A)$ and $Ker(A^*A) \subseteq Ker(A)$.
To prove the latter, suppose we have a vector $u \in Ker(A)$ so that $Au = 0$. Then $A^*Au = A^*(Au) = A^*0 = 0$, which means $u \in Ker(A^*A)$.
To prove the former, suppose we have a vector $v \in Ker(A^*A)$. We want to show that $Av = 0$. One way of

doing so is to show that its norm is 0.

$$\|Av\|^2 = (Av, Av) = (A^*v^*, A^*v^*) = A^*(v^*, A^*v^*) = A^*(Av, v) = (A^*Av, v) = (0, v) = 0$$

■

5.5 Adjoint of a Linear Transformation

Recall that the *Hermitian adjoint* A^* of matrix A is defined as the complex conjugate of each entry in A^T.

Theorem 5.5.1.

$$(Ax, y) = (x, A^*y) \qquad \forall x \in \mathbb{C}^n, \forall y \in \mathbb{C}^m$$

Proof.

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

The second equality uses the fact that because the adjoint consists of a transpose, we have $(AB)^* = B^*A^*$ and $(A^*)^* = A$. ■

This identity is used to define the adjoint operator.

Lemma 5.5.2. *The adjoint is unique.*

Proof. Suppose B satisfies $(Ax, y) = (x, By) \qquad \forall x, y$, then we can write

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

which means $A^* = B$. ■

Properties of the adjoint operator:

1. $(A + B)^* = A^* + B^*$
2. $(\alpha A)^* = \overline{\alpha} A^*$
3. $(AB)^* = B^*A^*$
4. $(A^*)^* = A$
5. $(y, Ax) = (A^*y, x)$

Theorem 5.5.3 (Relation between fundamental subspaces)**.** *Let $A : V \to W$ be an operator acting from one inner product space to another. Then*

1. $Ker(A^*) = (Range(A))^\perp$
2. $Ker(A) = (Range(A^*))^\perp$
3. $Range(A) = (Ker(A^*))^\perp$
4. $Range(A^*) = (Ker(A))^\perp$

Note that earlier we defined the fundamental subspaces using A^T instead of A^* because when discussing only $\mathbb{R}$ there was no difference.

Proof. Note that statements 1/3 and 2/4 are equivalent because for any subspace E, we have $(E^\perp)^\perp = E$. Also note that statement 2 is exactly statement 1 applied to the operator A^* since $(A^*)^* = A$.
Thus we only need to prove statement 1.
A vector $x \in (Range(A))^\perp$ means that x is orthogonal to all vectors of the form Ay, that is

$$(x, Ay) = 0 \qquad \forall y$$

Since $(x, Ay) = (A^*x, y)$, this is equivalent to

$$(A^*x, y) = 0 \qquad \forall y$$

This means that $A^*x = 0$, which means $x \in Ker(A^*)$. ■

We can use the idea of adjoint operators to prove another property of orthogonal projections.

Theorem 5.5.4. *If $P : V \to E$ is an orthogonal projection, it is self-adjoint.*

Proof. Recall that any v_k can be written as a sum of vectors in E and $E^\perp$, $v_k = e_k + \widetilde{e_k}$. Then

$$(Pv_1, v_2) = (u_1, u_2 + \widetilde{u_2}) = (u_1, u_2) + (u_1, \widetilde{u_2}) = (u_1, u_2)$$

and

$$(v_1, Pv_2) = (u_1 + \widetilde{u_1}, u_2) = (u_1, u_2) + (\widetilde{u_1}, u_2) = (u_1, u_2)$$

Since $(Tv_1, v_2) = (v_1, Tv_2)$, $T = T^*$. ■

5.6 Isometries and Unitary Operators

Definition 5.6.1. An operator $U : X \to Y$ is called an **isometry** if it preserves the norm,

$$\|Ux\| = \|x\| \qquad \forall x \in X$$

Theorem 5.6.1. *An operator $U : X \to Y$ is an isometry if and only if it preserves the inner product, ie if and only if*

$$(x, y) = (Ux, Uy) \qquad \forall x, y \in X$$

Proof. We use the polarization identities previously described. If X is a complex space

$$\begin{aligned}(Ux, Uy) &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \left\| Ux + \alpha Uy \right\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \left\| U(x + \alpha y) \right\|^2 \\ &= \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \left\| x + \alpha y \right\|^2 = (x, y)\end{aligned}$$

If X is a real space

$$\begin{aligned}(Ux, Uy) &= \frac{1}{4}(\|Ux+Uy\|^2 - \|Ux-Uy\|^2) \\ &= \frac{1}{4}(\|U(x+y)\|^2 - \|U(x-y)\|^2) \\ &= \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) = (x,y)\end{aligned}$$

■

Lemma 5.6.2. *An operator $U : X \to Y$ is an isometry if and only if $U^*U = I$.*

Proof. If $U^*U = I$, then

$$(x,x) = (U^*Ux, x) = (Ux, Ux) \qquad \forall x \in X$$

Since $\|x\| = \|Ux\|$, U is an isometry.
If U is an isometry, then by the above theorem and definition of adjoint

$$(U^*Ux, y) = (Ux, Uy) = (x, y) \qquad \forall x, y \in X$$

which means $U^*U = I$. ■

This lemma implies that an isometry is always left invertible since $U^*U = I$.

Definition 5.6.2. An isometry $U : X \to Y$ is called a **unitary operator** if it is invertible.

Lemma 5.6.3. *An isometry $U : X \to Y$ is a unitary operator if and only if $dim(X) = dim(Y)$.*

Proof. If $dim(X) = dim(Y)$, then U is square. Since we know U is left invertible, it must also then be invertible.
If U is unitary, it is invertible, so $dim(X) = dim(Y)$ since only square matrices are invertible. ■

Theorem 5.6.4. *The product of unitary operators is also unitary.*

Proof. Let A and B be unitary operators. We must show that $(AB)^*(AB) = I$.

$$(AB)^*(AB) = B^*A^*AB = B^*B = I$$

■

Properties of unitary operators that follow from our proofs:

1. $U^{-1} = U^*$
2. If U is unitary, $U^* = U^{-1}$ is also unitary.
3. If U is an isometry and $v_1, \cdots, v_n$ is an orthonormal basis, then $Uv_1, \cdots, Uv_n$ is an orthonormal basis.
4. The product of unitary operators is a unitary operator.

Lemma 5.6.5.

$$det(A^*) = \overline{det(A)}$$

Proof. Recall that the determinant of a matrix is equal to the product of its eigenvalues. We will show that for any eigenvalue λ of A, $\overline{\lambda}$ is an eigenvalue of A^*.
Note that λ is **not** an eigenvalue of A if and only if $A - \lambda I$ is invertible, which happens if and only if there exists an operator B such that

$$B(A - \lambda I) = (A - \lambda I)B = I$$

Taking the adjoints of all three sides means the above is equivalent to

$$(A^* - \overline{\lambda} I)B^* = B^*(A^* - \overline{\lambda} I) = I$$

Thus $A - \lambda I$ is invertible if and only if $A^* - \overline{\lambda} I$ is invertible, which means if λ is an eigenvalue of A, $\overline{\lambda}$ is an eigenvalue of A^*. ■

Theorem 5.6.6. *If U is a unitary matrix, then*

$$det(U) = \pm 1$$

If λ is an eigenvalue of U, then

$$\lambda = e^{i\theta}$$

for some real θ.

Proof. Let $det(U) = z$. Since $det(U^*) = \overline{det(U)}$, we have

$$|z|^2 = \overline{z}z = det(U^*U) = det(I) = 1$$

To prove statement 2, notice that the condition $\lambda = e^{i\theta}$ requires we prove $|\lambda| = 1$.
Now, if $Ux = \lambda x$, then

$$\|Ux\| = \|\lambda x\| = |\lambda| \cdot \|x\|$$

which means $|\lambda| = 1$ since $\|Ux\| = \|x\|$. ■

Definition 5.6.3. Operators A and B are called **unitarily equivalent** if there exists a unitary operator U such that $A = UBU^*$. Since for any unitary U, we have $U^{-1} = U^*$, any two unitarily equivalent matrices are similar as well.
The converse is **not** true.

The following theorem gives a way to construct a counter example to prove similar matrices are not always unitarily equivalent.

Theorem 5.6.7. *A matrix A is unitarily equivalent to a diagonal one if and only if it has an orthogonal (orthonormal) basis of eigenvectors.*

Proof. Using diagonalization, we can write $A = UBU^*$ and let $Bx = \lambda x$. Then $AUx = UBx = \lambda Ux$, which means Ux is an eigenvector of A.
Only if: Let A be unitarily equivalent to a diagonal matrix D, ie $A = UDU^*$. Because D is diagonal, the vectors e_k of the standard basis are eigenvectors of D, so Ue_k are eigenvectors of A. Since U is unitary,

$Ue_1, \cdots, Ue_n$ is an orthonormal basis.

If: Let A have an orthogonal basis $u_1, \cdots, u_n$ of eigenvectors. By dividing each vector by its norm, we can assure we have an orthonormal basis. By letting D be the matrix A in the basis $u_1, \cdots, u_n$, we know D will be a diagonal matrix.

By setting U to be the matrix with columns $u_1, \cdots, u_n$, we know U is unitary since its columns form an orthonormal basis (orthogonality implies invertibility and normality implies norm preservation). The change of coordinate formula implies

$$A = [A]_{SS} = [I]_{SB}[A]_{BB}[I]_{BS} = UDU^{-1} = UDU^*$$

where the last step follows from $U^{-1} = U^*$ for unitary matrices. ■

Chapter 6

Structure of Operators: WIP

Chapter 7

Outer Product Representation

This chapter is largely adapted from Chapter 2 of Nielsen and Chuang. Much of it will be spent providing an additional perspective on the core ideas in the previous two chapters. New concepts, like tensor products, that are introduced in the second half of this chapter will be developed more formally later in the book.

7.1 Dirac Notation

We can use inner products to derive a useful representations of linear operators. Before doing so, we'll define a convenient notation.

Definition 7.1.1. The **Dirac notation** represents vectors as

$$|\psi\rangle$$

which is also known as a **ket**. The dual of this same vector (dual spaces will be defined later in the book) is represented as

$$\langle\psi|$$

which is also known as a **bra**. For now, it suffices to know that the dual of a vector is simply the complex conjugate transpose of that vector. This notations allows us to represent the inner product of two vectors as

$$\langle\phi|\psi\rangle$$

Using this new notation, we can define the outer product.

Definition 7.1.2. The **outer product representation** is a representation of linear operators which uses the inner product. Suppose $|v\rangle$ is a vector in inner product space V, and $|w\rangle$ is a vector in inner product space W. We define $|w\rangle\langle v|$ to be the *linear operator* from V to W defined by

$$\Big(|w\rangle\langle v|\Big)|v_1\rangle \equiv |w\rangle\langle v|v_1\rangle = \langle v|v_1\rangle|w\rangle$$

To make this representation feel less abstract, we consider an important result using the outer product.

Lemma 7.1.1 (Completeness Relation)**.** *For an orthonormal basis $|i\rangle$,*

$$\sum_i |i\rangle\langle i| = I$$

Proof. We know that for any $v \in V$ can be written as $v = \sum_i v_i |i\rangle$ and that $\langle i|v\rangle = v_i$. Then

$$\left(\sum_i |i\rangle\langle i|\right)|v\rangle = \sum_i |i\rangle\langle i|v\rangle = \sum_i v_i |i\rangle = |v\rangle$$

Because this is true for all $|v\rangle$, it must be the identity operator. ∎

One application of the Completeness Relation is the representation of any operator in the outer product notation. Suppose $A : V \to W$ is a linear operator and $|v_i\rangle$ and $|w_j\rangle$ are orthonormal bases for V and W, respectively. By using the Completeness Relation twice, we get

$$A = I_W A I_V = \sum_{ij} |w_j\rangle\langle w_j|A|v_i\rangle\langle v_i| = \sum_{ij} \langle w_j|A|v_i\rangle |w_j\rangle\langle v_i|$$

which is the **outer product representation of** A. This equation also shows that A has matrix element $\langle w_j|A|v_i\rangle$ in the ith column and jth row, with respect to input basis $|v_i\rangle$ and output basis $|w_j\rangle$.
To gain more familiarity with the outer product representation, we will prove the Cauchy-Schwarz Inequality.

Theorem 7.1.2 (Cauchy-Schwarz Inequality)**.** *For any two vectors $|v\rangle$ and $|w\rangle$,*

$$|\langle v|w\rangle|^2 \leq \langle v|v\rangle\langle w|w\rangle$$

Proof. We use Gram-Schmidt to obtain an orthonormal basis $|i\rangle$, such that the first member of the basis is $|w\rangle/\sqrt{\langle w|w\rangle}$. Then

$$\begin{aligned}\langle v|v\rangle\langle w|w\rangle &= \sum_i \langle v|i\rangle\langle i|v\rangle\langle w|w\rangle \\ &\geq \frac{\langle v|w\rangle\langle w|v\rangle}{\langle w|w\rangle}\langle w|w\rangle \\ &= \langle v|w\rangle\langle w|v\rangle = |\langle v|w\rangle|^2\end{aligned}$$

where the inequality in the second step follows from the fact that we only use the first member of our basis and thus drop some non-negative terms. ∎

7.2 Diagonalization

The spectral theorem for normal operators proved in the previous chapter ($N = UDU^\dagger$ where U is a unitary matrix of eigenvectors and D is a diagonal matrix of corresponding eigenvalues) states an operator is diagonalizable if and only if it is normal. We can also write the diagonal representation of a normal matrix in outer product representation.

Theorem 7.2.1. *A diagonal representation for a normal operator $N : V \to V$ is*

$$N = \sum_i^n \lambda_i |i\rangle\langle i|$$

where $|i\rangle$ are an orthonormal set of eigenvectors of V with corresponding eigenvalues λ_i.

Proof. Let $|e\rangle$ be an orthonormal basis of N. Since each basis vector $|e\rangle$ can be understood as providing a coordinate of $|v\rangle$ in that basis through $\langle e|v\rangle$, we know $\langle e|e\rangle$ is the 0 matrix except with a 1 on the eth diagonal entry. Using the spectral theorem, we get

$$N = UDU^\dagger = U\left(\sum_{i=1}^n \lambda_i |e\rangle\langle e|\right)U^\dagger = \sum_{i=1}^n \lambda_i U|e\rangle\left(U|e\rangle\right)^\dagger$$

Now we define $U|e\rangle$ to be the normalized eigenvectors $|i\rangle$ to get

$$N = \sum_{i=1}^{n} \lambda_i |i\rangle \langle i|$$

■

As an example, we will introduce the Pauli $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ matrix, which is significant in quantum computation, and write its diagonal representation.
The characteristic equation $det(A - \lambda I)$ determines that the eigenvalues for Z are 1 and -1. The corresponding eigenvectors are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which we will denote $|0\rangle$ and $|1\rangle$, respectively. This means our diagonal representation is

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = |0\rangle \langle 0| - |1\rangle \langle 1|$$

7.3 Adjoints and Hermitian Operators

Recall, if A is a linear operator on V, then there exists a unique linear operator $A^\dagger$ (synonymous with A^* in this book) on V such that for all vectors $|v\rangle, |w\rangle \in V$,

$$(|v\rangle, A|w\rangle) = (A^\dagger |v\rangle, |w\rangle)$$

This linear operator is the **adjoint** or **Hermitian conjugate** of A. Recall earlier that we also defined $|v\rangle^\dagger \equiv \langle v|$.

Definition 7.3.1. A mapping $f : V \to W$ between complex vector spaces is **antilinear** if

$$f(ax + by) = \overline{a} f(x) + \overline{b} f(y)$$

From Wikipedia: "antilinear maps occur in quantum mechanics in the study of time reversal and in spinor calculus."

Lemma 7.3.1 (Antilinearity of the adjoint)**.**

$$\left(\sum_i a_i A_i \right)^\dagger = \sum_i a_i^* A_i^\dagger$$

Proof. On the LHS: we know $(a+b)^* = a^* + b^*$. Thus, $(\sum_i a_i A_i)^\dagger = \sum_i (a_i A_i)^\dagger$.
On the RHS: $\sum_i a_i^* A_i^\dagger = (a_1 A_1)^\dagger + \cdots + (a_n A_n)^\dagger = \sum_i (a_i A_i)^\dagger$. ■

Recall, an operator A whose adjoint is itself is known as a **Hermitian** or **self-adjoint** operator.
An interesting usage of Hermitian operators that hint at their importance is the following theorem.

Theorem 7.3.2. *An arbitrary operator A can be decomposed into the sum $B + iC$ where B and C are Hermitian.*

Proof. Define $B = \frac{A + A^\dagger}{2}$ and $C = \frac{A - A^\dagger}{2i}$.
To prove B is Hermitian: $B^\dagger = \frac{A^\dagger + A}{2} = B$. To prove C is Hermitian: $C^\dagger = \frac{A^\dagger - A}{-2i} = C$.

Finally,

$$B + iC = \frac{A + A^\dagger}{2} + i\frac{A - A^\dagger}{2i} = \frac{A + A}{2} = A$$

■

Recall, *projector* operators are Hermitian operators. We can demonstrate this using outer product representation.

Suppose W is a k-dimensional vector *subspace* of the d-dimensional vector space V. We use Gram-Schmidt to construct an orthonormal basis $|1\rangle, \cdots, |d\rangle$ for V such that $|1\rangle, \cdots, |k\rangle$ is an orthonormal basis for W. Then, we define

$$P = \sum_{i=1}^{k} |i\rangle \langle i|$$

to be the projector onto subspace W. Notice that the Completeness Relation states that if $k = d$, $P = I$, which is exactly what a projector from $V \to V$ would be. This outer product representation of P makes it clear that P is Hermitian: since $(|v\rangle \langle v|)^\dagger = (|v\rangle \langle v|)$, antilinearity means

$$P^\dagger = \left(\sum_{i}^{k} |i\rangle \langle i| \right)^\dagger = \sum_{i}^{k} (|i\rangle \langle i|) = \sum_{i}^{k} |i\rangle \langle i| = P$$

The Completeness Relation also makes it clear that the orthogonal complement of P is defined as $Q = I - P$, and that Q is a projector onto the vector space spanned by $|k+1\rangle, \cdots, |d\rangle$.

Recall, an operator A is **normal** if $AA^\dagger = A^\dagger A$. Clearly, a Hermitian operator is also normal.

Recall, an operator U is **unitary** if $U^\dagger U = I$, so unitary operators are also normal. In the Chapter 5, we proved unitary operators preserve the norm, but we can generalize this to prove that they also preserve inner products.

Theorem 7.3.3. *Let U be unitary. Then*

$$(U|v\rangle, U|w\rangle) = (|v\rangle, |w\rangle)$$

Proof.

$$(U|v\rangle, U|w\rangle) = \langle v| U^\dagger U |w\rangle = \langle v| I |w\rangle = \langle v|w\rangle$$

■

This result allows an elegant outer product representation of U. Let $|v_i\rangle$ be an orthonormal basis. We define

$$|w_i\rangle = U|v_i\rangle$$

so that $|w_i\rangle$ is also an orthonormal basis because unitary operators preserve inner products. Right multiplying both sides by $\langle v_i|$ and using the Completeness Relation, we get

$$|w_i\rangle \langle v_i| = U|v_i\rangle \langle v_i| \to U = \sum_i |w_i\rangle \langle v_i|$$

Recall, an operator A such that $(|v\rangle, A|v\rangle)$ is real and non-negative is called a **positive operator** or positive semidefinite. If this same inner product is *strictly* positive for all $|v\rangle \neq 0$, then A is **positive definite**.

Theorem 7.3.4 (Hermiticity of positive operators)**.** *A positive operator A is Hermitian.*

Proof. Since we can write $A = B + iC$ where B and C are Hermitian, we can use definition of a positive operator to write

$$\begin{aligned}(|v\rangle, A|v\rangle) &= (|v\rangle, B|v\rangle + iC|v\rangle) \\ &= (|v\rangle, B|v\rangle) + (|v\rangle, iC|v\rangle) \\ &= (B|v\rangle, |v\rangle) + (-iC|v\rangle, |v\rangle)\end{aligned}$$

This implies that $A = B - iC$, so

$$A^\dagger = (B - iC)^\dagger = B^\dagger + iC^\dagger = B + iC = A$$

■

7.4 Tensor Products

Tensor products are probably the most unintuitive Linear Algebra concept I've come across thus far. This section is a very top-level overview of the subject; later chapters will cover this topic in more detail.

Definition 7.4.1. Suppose V and W are m and n dimensional vector spaces, respectively. Then $V \otimes W$ is the **tensor product** of these two spaces and is an mn dimensional vector space. The elements of $V \otimes W$ are linear combinations of **tensor products** $|v\rangle \otimes |w\rangle$. If $|i\rangle$ and $|j\rangle$ are orthonormal bases for spaces V and W, then $|i\rangle \otimes |j\rangle$ is a basis for $V \otimes W$.

The tensor product satisfies the following properties:

1. For scalar z, $|v\rangle \in V$, and $|w\rangle \in W$,

$$z(|v\rangle \otimes |w\rangle) = (z|v\rangle) \otimes |w\rangle = |v\rangle \otimes (z|w\rangle)$$

2. For $|v_1\rangle, |v_2\rangle \in V$ and $|w\rangle \in W$,

$$(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle$$

3. For $|v\rangle \in V$ and $|w_1\rangle, |w_2\rangle \in W$,

$$|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = |v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle$$

If A and B are linear operators on V and W, respectively, then we can define a linear operator $A \otimes B$ on $V \otimes W$ as

$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A|v\rangle \otimes B|w\rangle$$

We can extend this definition of $A \otimes B$ to ensure linearity of $A \otimes B$ so

$$(A \otimes B)\left(\sum_i \alpha_i |v_i\rangle \otimes |w_i\rangle\right) \equiv \sum_i \alpha_i A|v_i\rangle \otimes B|w_i\rangle$$

An arbitrary linear operator C mapping $V \otimes W$ to $V' \otimes W'$ can be represented as a linear combination of tensor products of operators mapping $A : V \to V'$ and $B : W \to W'$,

$$C = \sum_i c_i A_i \otimes B_i$$

We can use the inner products on V and W to define an inner product on $V \otimes W$,

$$\left(\sum_i \alpha_i |v_i\rangle \otimes |w_i\rangle, \sum_j \beta_j |v_j'\rangle \otimes |w_j'\rangle\right) \equiv \sum_{ij} \alpha_i^* b_j \langle v_i|v_j'\rangle \langle w_i|w_j'\rangle$$

To make all this less abstract, we consider a matrix representation known as the *Kronecker product*. Suppose A is an $m \times n$ matrix and B is a $p \times q$ matrix. Then

$$A \otimes B \equiv \overbrace{\begin{bmatrix} A_{11}B & A_{12}B & \cdots & A_{1n}B \\ A_{21}B & A_{22}B & \cdots & A_{2n}B \\ \vdots & \vdots & & \vdots \\ A_{m1}B & A_{m2}B & \cdots & A_{mn}B \end{bmatrix}}^{nq} \Bigg\} mp$$

where $A_{11}B$ denotes the $p \times q$ submatrix B multiplied by the constant A_{11}.
For example,

$$\begin{bmatrix}1\\2\end{bmatrix} \otimes \begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}1\times 2\\1\times 3\\2\times 2\\2\times 3\end{bmatrix} = \begin{bmatrix}2\\3\\4\\6\end{bmatrix}$$

and the tensor product of Pauli X and Y is

$$X \otimes Y = \begin{bmatrix}0\cdot Y & 1\cdot Y\\1\cdot Y & 0\cdot Y\end{bmatrix} = \begin{bmatrix}0 & 0 & 0 & -i\\0 & 0 & i & 0\\0 & -i & 0 & 0\\i & 0 & 0 & 0\end{bmatrix}$$

The notation $|\psi\rangle^{\otimes k}$ denotes $|\psi\rangle$ tensored with itself k times.
Do problems from Nielsen and Chuang.

7.5 Operator Functions

Given a function $f : C \to C$, we can define a corresponding matrix function. Let $A = \sum_a a|a\rangle\langle a|$, then

$$f(A) \equiv \sum_a f(a)|a\rangle\langle a|$$

We can use this construction to define the square root of a positive operator, the logarithm of a positive-definite operator, or the exponential of a normal operator.
One important matrix function is *trace*. Since we know trace is *cyclic* ($tr(AB) = tr(BA)$) then $tr(UAU^\dagger) = tr(U^\dagger UA) = tr(A)$.
Suppose $|\psi\rangle$ is a unit vector and A is an arbitrary operator. To evaluate $tr(A|v\rangle\langle v|)$ use Gram-Schmidt to extend $|\psi\rangle$ to an orthonormal basis $|i\rangle$ which includes $|\psi\rangle$ as the first element. Then

$$tr(A|\psi\rangle\langle\psi|) = \sum_i \langle i|A|\psi\rangle\langle\psi|i\rangle = \langle\psi|A|\psi\rangle$$

7.6 Commutator and Anti-commutator

Definition 7.6.1. The **commutator** between two operators A and B is defined to be

$$[A, B] \equiv AB - BA$$

If $[A, B] = 0$, then we say A **commutes** with B. The **anti-commutator** is defined to be

$$\{A, B\} \equiv AB + BA$$

We say A *anti-commutes* with B if $\{A, B\} = 0$.

Many important properties of pairs of operators can be deduced from their commutator or anti-commutator. A useful relation is the connection between the commutator and the property of being able to *simultaneously diagonalize* Hermitian operators A and B, that is, write $A = \sum_i a_i|i\rangle\langle i|, B = \sum_i b_i|i\rangle\langle i|$, where $|i\rangle$ is a common set of eigenvectors for A and B.

Theorem 7.6.1 (Simultaneous diagonalization theorem)**.** *Suppose A and B are Hermitian operators. Then $[A,B]=0$ if and only if there exists an orthonormal basis such that both A and B are diagonal with respect to that basis.*

Proof. Proving if is simple, since if $A=UA_1U^{-1}$ and $B=UB_1U^{-1}$, then clearly $AB-BA=0$.
To prove only if, let $|a,j\rangle$ be an orthonormal basis for the eigenspace V_a of A with eigenvalue a; the index j being used to label possible degeneracies. Because we know $AB=BA$, we can write

$$AB|a,j\rangle = BA|a,j\rangle = aB|a,j\rangle$$

which means $B|a,j\rangle$ is an element of the eigenspace V_a. Let P_a denote the projector onto V_a and define $B_a = P_aBP_a$. Since projectors are Hermitian, we know B_a is also Hermitian so it must have a spectral decomposition in terms of an orthonormal set of eigenvectors that span V_a. Let's call these eigenvectors $|a,b,k\rangle$, where a and b label the eigenvalues of A and B_a, and k indexes degeneracies. Note that $B|a,b,k\rangle$ is in V_a, so $B|a,b,k\rangle = P_aB|a,b,k\rangle$. Additionally, we know $P_a|a,b,k\rangle = |a,b,k\rangle$, so

$$B|a,b,k\rangle = P_aBP_a|a,b,k\rangle = b|a,b,k\rangle$$

This means $|a,b,k\rangle$ is an eigenvector of B with eigenvalue b, and therefore $|a,b,k\rangle$ is a spanning orthonormal set of eigenvectors of both A and B. Thus, A and B are simultaneously diagonalizable. ■

Another use of the commutator and anti-commutator is the following identity:

$$AB = \frac{[A,B]+\{A,B\}}{2}$$

Lemma 7.6.2. *Suppose $[A,B]=0, \{A,B\}=0$, and A is invertible. Then B is 0.*

Proof. Using the identity we just defined, we know $AB=\frac{0+0}{2}$. Since A is invertible we left multiply both sides by A^{-1} to get $B=0$. ■

Two more easily verifiable identities:

1. $[A,B]^\dagger = [B^\dagger, A^\dagger]$
2. $[A,B] = -[B,A]$
3. If A,B are Hermitian, then $i[A,B]$ is also Hermitian.

7.7 Polar Decomposition

We don't understand *general* linear operators incredibly well, but we do have a good understanding of *unitary* and *positive* operators. The polar and singular value decompositions allow us to take advantage of this by breaking general linear operators into products of unitary and positive operators.

Theorem 7.7.1 (Polar decomposition)**.** *Let A by a linear operator on V. Then there exists unitary U and positive operators J and K such that*

$$A = UJ = KU$$

where $J=\sqrt{A^\dagger A}$ and $K=\sqrt{AA^\dagger}$. Additionally, if A is invertible, then U is unique.

Proof. We know J (which is the modulus of A) is a positive operator (which we know are normal), so we can give it a spectral decomposition,

$$J = \sum_i \lambda_i |i\rangle \langle i| \qquad (\lambda_i \geq 0)$$

Define $|\psi_i\rangle = A|i\rangle$, so that $\langle \psi_i | \psi_i \rangle = \lambda_i^2$.
Now consider only the positive eigenvalues. With these eigenvalues, define $|e_i\rangle = |\psi_i\rangle / \lambda_i$, so the $|e_i\rangle$ are normalized and are orthogonal, since if $i \neq j$, then $\langle e_i | e_j \rangle = \langle i | A^\dagger A | j \rangle / \lambda_i \lambda_j = 0$.
Now we use Gram-Schmidt to extend the orthonormal set $|e_i\rangle$ to form an orthonormal basis, which we'll also call $|e_i\rangle$.
We define the unitary operator $U = \sum_i |e_i\rangle \langle i|$. When $\lambda_i \neq 0$, we have $UJ|i\rangle = \lambda_i |e_i\rangle = |\psi_i\rangle = A|i\rangle$. When $\lambda_i = 0$, we have $UJ|i\rangle = 0 = |\psi_i\rangle$. Since A and UJ agree on the basis $|i\rangle$, we know $A = UJ$.
J is unique since left-multiplying $A = UJ$ by $A^\dagger = JU^\dagger$ gives $A^\dagger A = J^2$, from which we get $J = \sqrt{A^\dagger A}$. The equation $A = UJ$ also means that if A is invertible, so is J, which means U is uniquely determined by $U = AJ^{-1}$.
Proving the right polar decomposition $A = KU$ follows since

$$A = UJ = UJU^\dagger U = KU$$

where $K = UJU^\dagger$ is a positive operator. And since $AA^\dagger = KUU^\dagger K = K^2$, we have $K = \sqrt{AA^\dagger}$. ∎

7.8 Singular Value Decomposition

The singular value decomposition combines polar decomposition and the spectral theorem.

Theorem 7.8.1 (Singular value decomposition)**.** *Let A be a linear operator on V. Then there exist unitary matrices U and V, and a diagonal matrix D with non-negative entries such that*

$$A = UDV$$

The diagonal entries of D are called the singular values of A.

Proof. By the polar decomposition, $A = SJ$, for unitary S and positive operator J. By the spectral theorem, $J = TDT^\dagger$, for unitary T and diagonal D with non-negative entries.
Now we just set $U = ST$ and $V = T^\dagger$ to get

$$A = SJ = STDT^\dagger = UDV$$

∎

Chapter 8

Matrix Functions - WIP

A **matrix function** is a function that maps one matrix to another.

8.1 Matrix Exponential

Most of this section is from the Wikipedia article on the matrix exponential and Dan Klain's notes available at

Definition 8.1.1. The **matrix exponential** is a matrix function on square matrices similar to the scalar exponential function. It is used to solve systems of linear differential equations.
Let X be an $n \times n$ matrix. The exponential is given by the power series,

$$e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k = I + X + \frac{1}{2!}X^2 + \frac{1}{3!}X^3 + \cdots$$

where X^0 is the identity matrix. Note that if X is a 1×1 matrix, the exponential of X is a 1×1 matrix whose element is the exponential of X's sole element.

Two properties that are clear from this definition are

1. $exp(0) = I$
2. $exp(X^*) = (exp(X))^*$

Another useful property is the following result.

Lemma 8.1.1. *If Y is invertible, then*

$$e^{YXY^{-1}} = Ye^XY^{-1}$$

Proof.

$$\begin{aligned} e^{YXY^{-1}} &= I + YXY^{-1} + \frac{1}{2!}(YXY^{-1})^2 + \frac{1}{3!}(YXY^{-1})^3 + \cdots \\ &= I + YXY^{-1} + Y\frac{X^2}{2!}Y^{-1} + Y\frac{X^3}{3!}Y^{-1} + \cdots \\ &= Y\left(I + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots\right)Y^{-1} \\ &= Ye^XY^{-1} \end{aligned}$$

■

Not all properties of the scalar exponential function translate to the matrix exponential. For example, we know $e^{a+b} = e^a e^b$ when a and b are scalars, but the matrix exponential version of this typically does not hold. There are a few exceptions, however, as the following lemma illustrates.

Lemma 8.1.2. *Let A be a square matrix, and let $s, t \in \mathbb{C}$. Then*

$$e^{A(s+t)} = e^{As}e^{At}$$

Proof.

$$\begin{aligned} e^{As}e^{At} &= \left(I + As + \frac{A^2s^2}{2!} + \cdots\right)\left(I + At + \frac{A^2t^2}{2!} + \cdots\right) \\ &= \left(\sum_{j=0}^{\infty} \frac{A^j s^j}{j!}\right)\left(\sum_{k=0}^{\infty} \frac{A^k t^k}{k!}\right) \\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{A^{j+k}s^j t^k}{j!k!}\Bigg) \end{aligned}$$

If we let $n = j + k$ so that $k = n - j$, we can write

$$e^{As}e^{At} = \sum_{j=0}^{\infty}\sum_{n=j}^{\infty} \frac{A^n s^j t^{n-j}}{j!(n-j)!} = \sum_{n=0}^{\infty} \frac{A^n}{n!}\sum_{j=0}^{n} \frac{n!}{j!(n-j)!} s^j t^{n-j} = \sum_{n=0}^{\infty} \frac{A^n(s+t)^n}{n!} = e^{A(s+t)}$$

where the second equality follows from rearranging terms and the third follows from the Binomial Theorem. ■

If we set $s = 1$ and $t = -1$, we get

$$e^A e^{-A} = e^{A(1-1)} = e^0 = I$$

which means that regardless of what matrix A is, the exponential matrix e^A is **always** invertible, with inverse e^{-A}.

An important motivation for the matrix exponential is its use in differential equations.

Theorem 8.1.3. *Let A be a square matrix, and let t be a real scalar. Let $f(t) = e^{tA}$. Then*

$$f'(t) = Ae^{tA}$$

Proof. Applying the previous lemma to the limit yields

$$f'(t) = \lim_{h\to 0} \frac{e^{A(t+h)} - e^{At}}{h} = e^{At}\left(\lim_{h\to 0} \frac{e^{Ah} - I}{h}\right)$$

Expanding $e^{Ah} - I$ gives us

$$f'(t) = e^{At}\left(\lim_{h\to 0} \frac{1}{h}\left[Ah + \frac{A^2h^2}{2!} + \cdots\right]\right) = e^{At}\left(\lim_{h\to 0} A + \frac{A^2h}{2!} + \cdots\right) = e^{At}A = Ae^{At}$$

■

Recall that if A and B are matrices, then usually $e^A e^B \neq e^{A+B}$. The following theorem provides a condition for this equality to hold.

Theorem 8.1.4. *Let A, B be $n \times n$ matrices. If A and B are commuting matrices ($AB = BA$), then*

$$e^{A+B} = e^A e^B$$

Proof. If $AB = BA$, we use the definition of the matrix exponential to write $Ae^{Bt} = e^{Bt}A$, and similarly for other combinations of A, B, $A + B$, and their exponentials.

Let $g(t) = e^{(A+B)t}e^{-Bt}e^{-At}$, where t is a real scalar. By the previous theorem and the product rule for

derivatives, we have

$$\begin{aligned} g'(t) &= (A+B)e^{(A+B)t}e^{-Bt}e^{-At} + e^{(A+B)t}(-B)e^{-Bt}e^{-At} + e^{(A+B)t}e^{-Bt}(-A)e^{-At} \\ &= (A+B)g(t) - Bg(t) - Ag(t) \\ &= 0 \end{aligned}$$

Note that we could only factor out $(-A)$ and $(-B)$ because we know $AB = BA$.
Since the derivative is 0 for all t, $g(t)$ must be a matrix of constants, so $g(t) = C$, for some constant matrix C. Setting $t = 0$ gives us

$$C = g(0) = e^{(A+B)0}e^{-B0}e^{-A0} = I$$

which must also be true for all t since C is constant so

$$e^{(A+B)t}e^{-Bt}e^{-At} = I$$

and by multiplying both sides by $e^{At}e^{Bt}$, we get

$$e^{At}e^{Bt} = e^{(A+B)t}$$

■