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, as well as Linear Algebra Done Wrong by Treil.

June 28, 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.

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

∎

1.4 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 $(AB)^T = B^TA^T$,

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

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 $A = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix}$ 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 $\left[A \mid I\right]$.
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 $\left[A \mid I\right] \to \left[EA \mid EI\right] = \left[I \mid A^{-1}\right]$.

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,

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

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 equation 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 $0 \le i \le k$, and since eigenvalues are distinct, $c_i = 0$ for $0 \le i \le k$.
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 this chapter.

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 $|z|^2 = x^2 + y^2 = z\overline{z}$, where $\overline{z}$ 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-Schwartz Inequality)**.**

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

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:

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

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:

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

Using property 1 of inner products, we have

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

Subbing in t, we get

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

which completes the proof. ∎

Theorem 5.1.5 (Triangle Inequality)**.**

$$\|x, y\| \le \|x\| + \|y\|$$

Proof.

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

∎

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

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

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

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

Proof. For the real case,

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

For the complex case,

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

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 so far 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*

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

Proof.

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

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

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