Linear Algebra 1

Labix

April 4, 2024

Abstract

Linear algebra is at the heart of mathematics. Almost all areas of mathematics make some use of the notion of vector spaces and its properties. It began with the study of systems of linear equations.

Nowadays as we understand vector spaces independently from systems of linear equations, we will also treat the material differently. The first three chapter begins with the basis definitions: vector spaces and the maps between them called linear maps. Eigenvalues and eigenspaces will be an important invariant for vector spaces. Together with the aid of matrices, we will have a good grasp of how to write a given vector space in a simpler form. Chapter 4 will then improve on the further simplifying a given matrix so that we can read information from each easily.

The rest of the chapters will focus on particular properties of vector spaces and linear maps. They each correspond to an important class of matrices. For examples, quadratic form corresponds to matrices equivalent up to congruency while orthogonality of basis vectors give orthogonal matrices.

Contents

1 Vector Spaces

1.1 Introduction to Vector Spaces

Definition 1.1.1: Vector Space

A vector space V over a field $\mathbb{F}$ is an abelian group V whose binary operation is called vector addition, together with a group action $\cdot : \mathbb{F} \times V \to V$ called scalar multiplication. In particular, this means that the following properties are satisfied.
$(V, +)$ is an abelian group.

- $\mathbf{u} + \mathbf{v} \in V$
- $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- There exists a vector $\mathbf{0}_V$ such that $\mathbf{0}_V + \mathbf{u} = \mathbf{u}$
- There exists an additive inverse $-\mathbf{u}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

$\mathbb{F}$ acts on V as a group action, with an identity in V.

- $a \cdot \mathbf{u} \in V$
- $a(b\mathbf{u}) = (ab)\mathbf{u}$
- There exists a vector 1_V such that $1_v \cdot \mathbf{u} = \mathbf{u}$

Distributive laws.

- $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
- $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$

Proposition 1.1.2

Let $a \in \mathbb{F}$ and $u \in V$ be a vector space over $\mathbb{F}$.

- $a \cdot \mathbf{0}_V = \mathbf{0}_V$
- $0 \cdot \mathbf{u} = \mathbf{0}_V$
- $(-a)\mathbf{v} = -(a\mathbf{v}) = a(-\mathbf{v})$
- $a\mathbf{v} = \mathbf{0}_V \implies a = 0$ or $\mathbf{v} = \mathbf{0}_V$

Proof.

- $a(\mathbf{0}_V) = a(\mathbf{0}_V + \mathbf{0}_V) = a\mathbf{0}_V + a\mathbf{0}_V$. Adding the additive inverse of $a\mathbf{0}_V$ on both sides gives our result.
- $0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}$. Adding the additive inverse of $0\mathbf{u}$ on both sides gives our result.
- Naturally $-(a\mathbf{v})$ is the inverse of $a\mathbf{v}$. Consider $a\mathbf{v} + a(-\mathbf{v})$.
  $a\mathbf{v} + a(-\mathbf{v}) = a(\mathbf{v} - \mathbf{v}) = a\mathbf{0}_V = 0_V$. Thus $a(-\mathbf{v})$ is also the inverse of $a\mathbf{v}$ and $-(a\mathbf{v}) = a(-\mathbf{v})$. The same could be done to the third item with the other distributive law.
- Suppose that $a \neq 0$. Then $\mathbf{v} = (a^{-1}a)\mathbf{v} = a^{-1}(a\mathbf{v}) = 0$.

$\square$

Proposition 1.1.3

The additive identity, multiplicative identity, additive inverse of a vector space is unique.

Proof. Suppose that $\mathbf{e}$ and $\mathbf{f}$ are additive identities. Then $\mathbf{e}+\mathbf{f}=\mathbf{e}$ and $\mathbf{e}+\mathbf{f}=\mathbf{f}$. Thus $\mathbf{e}=\mathbf{f}$. Suppose that $\mathbf{e}$ and $\mathbf{f}$ are multiplicative identities. Then $\mathbf{ef}=\mathbf{e}$ and $\mathbf{ef}=\mathbf{f}$ and $\mathbf{e}=\mathbf{f}$. Let $a \in V$. Suppose that $b, c \in V$ are additive inverses of a. Then

$$\begin{aligned} a+b=a+c &\implies b+a+b=b+a+c \\ &\implies (b+a)+b=(b+a)+c \\ &\implies b=c \end{aligned}$$

□

1.2 Basis and Dimension

Definition 1.2.1: Linearly Independent

We say that a set of vectors $\{v_1, \dots, v_n\}$ of a vector space V over $\mathbb{F}$ are linearly independent if

$$\sum_{k=1}^n a_k v_k = 0$$

for $a_1, \dots, a_n \in \mathbb{F}$ implies $a_1 = \cdots = a_n = 0$.

Definition 1.2.2: Span

We say that a set of vectors $\{v_1, \dots, v_n\}$ of a vector space V over $\mathbb{F}$ spans V if for all $v \in V$, there exists $a_1, \dots, a_n \in \mathbb{F}$ such that

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

Definition 1.2.3: Basis

We say that a set of vectors $\{v_1, \dots, v_n\}$ of a vector space forms a basis for V if they are linearly independent and spans V.

Definition 1.2.4: Dimension

We say that the dimension of a vector space V is the number of elements in a basis of V. If a basis has n elements, then we say that $\dim(V) = n$.

If n is a finite number, then we say that V is finite dimensional.

We have yet to shown that the dimension of a vector space is well defined since we do not know whether the cardinality of any two bases are the same. Therefore we have the following important theorem for finite dimensional vector space.

Theorem 1.2.5: Steinitz Exchange Lemma

Let U, W be finite subsets of a finite dimensional vector space V. If U is a set of linearly independent vectors and W spans V, then

- $|U| \leq |W|$
- There exists a set $W' \subset W$ with $|W'| = |W| - |U|$ such that $U \cup W'$ spans V.

Proof. Take $U = \{u_1, \dots, u_m\}$ and $W = \{w_1, \dots, w_n\}$. We will show that after reordering elements of W, we will have a set $\{u_1, \dots, u_m, w_m + 1, \dots, w_n\}$ that it spans V. We proceed by induction on m. Suppose that $m = 0$. In this case, $|U| \leq |W|$ necesssarily holds and by construction, W already spans V.

Now suppose that the proposition is true for $m - 1$. By the induction hypothesis, we may reorder elements of W so that $\{u_1, \dots, u_{m-1}, w_m, \dots, w_n\}$ spans V. Since $u_m \in V$, there exists $a_1, \dots, a_n$ such that

$$u_m = \sum_{k=1}^{m-1} a_k u_k + \sum_{k=m}^{n} a_k w_k$$

At least one of $a_m, \dots, a_n$ must be nonzero else the equality will contradict the linear independence of $u_1, \dots, u_m$. This must mean that $m \leq n$.

Now by reordering $a_m w_m, \dots, a_n w_n$, we may assume that $a_m \neq 0$. Thus we have that

$$w_m = \frac{1}{a_m}\left(u_m - \sum_{k=1}^{m-1} a_k u_k - \sum_{k=m+1}^{n} a_k w_k\right)$$

This means that w_m lies in the span of $\{u_1, \dots, u_m, w_{m+1}, \dots, w_n\}$. Since this span contains each of the vectors $u_1, \dots, u_{m-1}, w_m, \dots, w_n$, by the inductive hypothesis it spans V. □

Clearly this implies that linearly independent sets of vectors must have cardinality less than sets of vectors that span V. By taking the highest cardinality of such linearly indendent set of vectors, and the lowest cardinality of such sets of vectors that span V, we necessarily have that they are equal and thus is exactly the dimension of V.

We will discuss about dimensions and infinite dimensional vector spaces more in functional analysis. For the rest of the notes we will mostly go with finite dimensional vector spaces.

We now give a criterion with matrices to find whether a set of vectors span V or whether they are linearly independent.

Theorem 1.2.6

Let V be a vector space of dimension n and $S = \{v_1, \dots, v_n\} \subset V$. Then

- Elements of S are linearly independent if and only if the row echelon form of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ has a leading one in every column
- Elements of S span V if and only if the row echelon form of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ has no zero rows
- S is a basis of V if and only if the row echelon form of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ is equal to the identity

1.3 Vector Subspaces

Definition 1.3.1: Vector Subspaces

A subset U of a vector space V is called a subspace of V if U is also a vector space.

Proposition 1.3.2: Subspace Criterion

U is a subspace of V if and only if U is closed under vector addition and scalar multiplication and contains the zero vector.

Proof. Suppose that U is a subspace of V. Then necessarily U is closed under vector addition and scalar multiplication and contains the zero vector.

Now suppose that the latter conditions are fulfilled by a subset U of V. Then it is easy to see that U satisfies all the criteria for being a vector space. □

Proposition 1.3.3

If U_1 and U_2 are subspaces of V then $U_1 \cap U_2$ is also a subspace.

Proof. Suppose that $\mathbf{v}, \mathbf{w} \in U_1 \cap U_2$. Then $\mathbf{v}, \mathbf{w} \in U_1$ and U_2. Since U_1, U_2 are subspaces, $\mathbf{v} + \mathbf{w} \in U_1$ and U_2 thus $\mathbf{v} + \mathbf{w} \in U_1 \cap U_2$. The proof is similar for scalar multiplication. □

Definition 1.3.4: Sum of Subspaces

Let U, W be subspaces of the vector space V. Then define

$$U + W = \{\mathbf{u} + \mathbf{w} : \mathbf{u} \in U \text{ and } \mathbf{w} \in W\}$$

Proposition 1.3.5

Let U, W be subspaces of a vector space V. Then $U + W$ is the smallest subspace of V containing U and W.

Proof. We first show that $U + W$ is indeed a subspace of V. Suppose that $v \in U + W$. Then there exists $u \in U$ and $w \in W$ such that $v = u + w$. Then since U and W are closed individually under vector addition and scalar multiplication, any product and addition in $U + W$ can be decomposed into a sum of vectors in U and W and thus the new vector will also be able to be decomposed into U and W and thus lie in $U + W$.

Now suppose that S is a subspace of V containing U and W. This means that any linear combination of elements of U and W are contained in S thus $U + W \subseteq S$. This means that if any subspace containing U and W must also contain $U + W$, which means that $U + W$ is the smallest subspace containing U and W. □

Definition 1.3.6: Independent Subspaces

Let $W_1, \dots, W_n$ be subspaces of a vector space V. We say that $W_1, \dots, W_n$ are independent if no vector of W_i is a linear combination of the remaining subspaces for every $i \in \{1, \dots, n\}$

Definition 1.3.7: Direct Sum

A vector space is the direct sum of its subspaces

$$V = W_1 \oplus \cdots \oplus W_n$$

if $W_1, \dots, W_n$ are independent and $V = W_1 + \cdots + W_n$.

Corollary 1.3.8

If $V = W_1 \oplus \cdots \oplus W_n$ then
$$\dim(V) = \sum_{k=1}^{n} \dim(W_k)$$

Proof. Each basis of W_k are not contained in any other linear combination of all the basis of $W_1, \dots, W_{k-1}, W_{k+1}, \dots, W_n$. This means that the set of all the basis of $W_1, \dots, W_n$ are linearly independent. Since they each span W_k independently, the set of all the basis of $W_1, \dots, W_n$ will span $W_1 \oplus \cdots \oplus W_n$ and thus is a basis of V. Thus we are done. □

1.4 Row and Column Ranks

The final section is devoted to matrices as we will soon see that matrices are particularly useful in a lot of things.

Definition 1.4.1: Row Space

Let $A_{m\times n}$ be a matrix. The row space of A is the subspace of $\mathbb{F}^m$,
$$\text{span}\{r_1, \dots, r_m\}$$
where r_i are the rows of A. The row rank of A is defined to be the dimension of the row space of A.

Definition 1.4.2: Column Space

Let $A_{m\times n}$ be a matrix. The column space of A is the subspace of $\mathbb{F}^n$,
$$\text{span}\{c_1, \dots, c_m\}$$
where c_i are the columns of A. The column rank of A is defined to be the dimension of the column space of A.

Lemma 1.4.3

Applying row operations does not change the row space, row rank of a matrix and column rank of a matrix.

Theorem 1.4.4

The row rank of a matrix is equal to the column rank.

We can now define the rank of a matrix without problem.

Definition 1.4.5: Rank of a Matrix

Define the rank of a matrix to be its row rank or column rank.

Proposition 1.4.6

Let A be a $n \times n$ matrix. Then the following are equivalent.

- The rank of A is n
- A is invertible
- The rows of A form a linearly independent set

- The columns of A form a linearly independent set

2 Linear Maps

2.1 Properties of Linear Maps

Definition 2.1.1: Linear Transformation

Let V, W be vector spaces over $\mathbb{F}$. A linear transformation or linear map T from V to W is a function $T : V \to W$ such that

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$
- $T(kv) = kT(v)$ for all $k \in \mathbb{F}$, $v \in V$

Lemma 2.1.2

Let $T : V \to W$ be a linear map.

- $T(0_v) = 0_w$
- $T(-v) = -T(v)$ for all $v \in V$

Proof. Suppose that $v \in V$. Then $T(0 \cdot v) = 0 \cdot T(v) = 0$. Also we have that $T(0 - v) = T(0) - T(v) = -T(v)$. □

Proposition 2.1.3

If $T : U \to V$ and $S : V \to W$ are linear then $S \circ T : U \to W$ is also linear.

Proof. Let $au + bv \in U$.

$$\begin{aligned} S \circ T(au + bv) &= S(aT(u) + bT(v)) \\ &= a(S \circ T(u)) + b(S \circ T(v)) \end{aligned}$$

□

2.2 Isomorphisms

Definition 2.2.1: Isomorphic Linear Maps

A linear map $T : V \to W$ is said to be an isomorphism if T is bijective. In this case we also say that V and W are isomorphic.

Theorem 2.2.2

Let $T : V \to W$ be an isomorphism of vector spaces V, W over F. Then its inverse map $T^{-1} : W \to V$ is a linear map.

Theorem 2.2.3

Let $T : V \to W$ be a linear map. Then the following are equivalent.

- T is isomorphic
- If $v_1, \dots, v_n \in V$ is a basis of V then $T(v_1), \dots, T(v_n) \in W$ is a basis of W

Corollary 2.2.4

Every finite dimensional vector space is isomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N} \setminus \{0\}$.

Proof. Direct consequence from the above. □

This corollary is especially important since it tells use that we only really need to study all of $\mathbb{R}^n$ to study all of finite dimensional spaces. Once we have our results on $\mathbb{R}^n$, we can translate it via an isomorphism.

Proposition 2.2.5

Let V, W be vetor spaces. The set of all linear maps from V to W forms a vector space. Denote it as $\mathcal{L}(V, W)$

2.3 Kernels and Images

Definition 2.3.1: Images and Kernels

Let $T : U \to V$ be a linear map. The image of T is defined as

$$\text{im}(T) = \{\mathbf{v} \in V | T(\mathbf{u}) = \mathbf{v}, \forall \mathbf{u} \in U\}$$

The kernel of T is defined as

$$\ker(T) = \{\mathbf{u} \in U | T(\mathbf{u}) = \mathbf{0}_V\}$$

Theorem 2.3.2

Let $T : U \to V$ be a linear map. Then

- $\text{im}(T)$ is a subspace of V
- $\ker(T)$ is a subspace of U

Proof. Let $T : U \to V$ be a linear map.

- We prove that $\text{im}(T)$ is a subspace of V. Let $\mathbf{u}, \mathbf{v} \in \text{im}(T)$ and $a \in \mathbb{F}$. Since $\mathbf{u}, \mathbf{v} \in \text{im}(T)$, there exists $\mathbf{u}_0, \mathbf{v}_0 \in U$ such that $T(\mathbf{u}_0) = \mathbf{u}$ and $T(\mathbf{v}_0) = \mathbf{v}$. Note that $a\mathbf{u}_0 \in U$ and $\mathbf{u}_0 + \mathbf{v}_0 \in U$. Consider $T(a\mathbf{u}_0)$. We have $T(a\mathbf{u}_0) = aT(\mathbf{u}_0) = a\mathbf{u}$. Thus $a\mathbf{u} \in \text{im}(T)$. Similarly, $T(\mathbf{u}_0 + \mathbf{v}_0) = T(\mathbf{u}_0) + T(\mathbf{v}_0) = \mathbf{u} + \mathbf{v}$. Thus $\mathbf{u} + \mathbf{v} \in \text{im}(T)$. By the subspace criterion $\text{im}(T)$ is a subspace of V.
- We now prove that $\ker(T)$ is a subspace of U. Suppose that $u, v \in \ker(T)$ and $a, b \in \mathbb{F}$. Then $T(au + bv) = aT(u) + bT(v) = 0$. Thus $au + bv \in \ker(T)$.

□

Definition 2.3.3: Rank and Nullity

Let $T : U \to V$ be a linear map.

- $\text{rank}(T) = \dim(\text{im}(T))$ is said to be the rank of T.
- $\text{nullity}(T) = \dim(\ker(T))$ is said to be the nullity of T.

Theorem 2.3.4: Rank Nullity Theorem

Let $T : U \to V$ be a linear map. Then

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

Theorem 2.3.5

Let $T : U \to V$ be a linear map, where $\dim(U) = n$, $\dim(V) = m$. Let $e_1, \dots, e_n$ be a basis of U. Then the rank of T is equal to the largest size of a linearly independent subset of $T(e_1), \dots, T(e_n)$.

2.4 Role of Matrices

Definition 2.4.1: Matrix of a Linear Map

Let $T : U \to V$ be a linear map where $\dim(U) = n$ and $\dim(V) = m$. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the standard basis of U and $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ the standard basis of V. Let

$$T(\mathbf{e}_i) = \sum_{k=1}^{m} \alpha_{ki}\mathbf{f}_k$$

for $i \in \{1, \dots, n\}$ Define the matrix of this linear map to be

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

Theorem 2.4.2

Let $T : U \to V$ be a linear map. Let A be the matrix of a linear map. Let $\mathbf{v} \in U$. Then the coordinates of $T(\mathbf{v})$ are given by

$$T(\mathbf{v}) = A\mathbf{v}$$

Theorem 2.4.3

The rank of a matrix equals the rank of any map that it represents.

Theorem 2.4.4

The composition of linear maps is represented by the matrix product of its representatives.

Theorem 2.4.5

Let $T : V \to W$ be a linear map. Then T is a vector space isomorphism if and only if its matrix is nonsingular.

3 Eigenspaces

Eigenspaces are invariants of a linear map. Every vector in the eigenspace will only be scaled while maintaining its direction.

3.1 Eigenvalues and Eigenvectors

Definition 3.1.1: Eigenvalues and Eigenvectors

Let $T : V \to V$ be a linear map, where V is a vector space over $\mathbb{F}$. Suppose that for some non-zero vector $v \in V$, and some scalar $\lambda \in \mathbb{F}$, we have $T(v) = \lambda v$. Then v is called an eigenvector of T, and λ is called the eigenvalue of T.

Notice that λ can in fact be 0. If this is the case, then the eigenvectors are just the vectors in the kernel.

Definition 3.1.2: Eigenspace

Let λ be an eigenvalue of a linear map T. The set of all eigenvectors belonging to λ is called an eigenspace of T with respect to λ, denoted E_λ.

Lemma 3.1.3

Let λ be an eigenvalue of A. Then
$$E_\lambda = \ker(A - \lambda I)$$

Proof. Clearly since $Av = \lambda v$ for any eigenvector v of λ, we also have that $(A - \lambda I)v = 0$ which means that $v \in \ker(A - \lambda I)$. □

Proposition 3.1.4

Let $\lambda_1, \dots, \lambda_r$ be distinct eigenvalues of $T : V \to V$, and let $v_1, \dots, v_r$ be the corresponding eigenvectors. Then $v_1, \dots, v_r$ are linearly independent.

As we can see, distinct eigenvalues are linearly independent. Considering the span of each eigenvectors, we can clearly see that each of their spans are independent.

Proposition 3.1.5

If $\lambda_1, \dots, \lambda_n$ are distinct eigenvalues of a matrix A, then $E_{\lambda_1}, \dots, E_{\lambda_n}$ are independent.

Proof. Clear from the fact that the basis of eigenspaces of different eigenvalues are linearly independent. □

Definition 3.1.6: Characteristic Polynomial

Let A be an $n \times n$ matrix.
$$c_A(x) = \det(A - xI_n)$$
is called the characteristic polynomial of A.

Proposition 3.1.7

Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if
$$c_A(\lambda) = 0$$

Definition 3.1.8: Invariant Subspaces

Let $T : V \to V$ be a linear transformation. Let U be a subspace of V. We say that U is T-invariant if

$$v \in U \implies T(v) \in U$$

for all $v \in U$ or equivalently, $T(U) \subseteq U$.

Theorem 3.1.9

Eigenspaces is an invariant subspace under its linear transformation.

The main result of this subsection, stated that eigenspaces remain invariant under the linear transformation. Clearly this depends on the linear transformation. We will also show that this fact is also unchanged when considering different basis for the linear transformation.

3.2 Change of Basis

Definition 3.2.1: Change of Basis Matrix

Let V be a vector space and B, B' are two basis of V. A change of basis linear map is a linear map $T : V \to V$ such that $T : V_B \to V_{B'}$, meaning the old basis is mapped to the new basis.

Proposition 3.2.2

Let V be a vector space and $v_1, \dots, v_n$, $v_1', \dots, v_n'$ two distinct basis of V. Then

$$v_k = p_{k1}v_1' + \cdots + p_{kn}v_n'$$

for all $k \in \{1, \dots, n\}$ and for any vector x in the basis $v_1, \dots, v_n$, the vector in the other basis x' is given by $x' = Px$ with the invertible matrix P

$$\begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

Theorem 3.2.3

Let V, W be vector spaces. Let V consists of two different basis B and B' with a map $P : V_B \to V_{B'}$. Similarly for W we have C and C' and $Q : W_C \to W_{C'}$. Suppose $A : V_B \to W_C$ is a linear map. Then $A' : V_{B'} \to W_{C'}$ is given by

$$A' = QAP^{-1}$$

Definition 3.2.4: Similar Matrices

We say that two matrices $A, B \in M_{n\times n}(\mathbb{R})$ are similar if there exists an invertible matrix $P \in M_{n\times n}(\mathbb{R})$ such that $B = PAP^{-1}$.

Lemma 3.2.5

The relation of similarity in matrices is an equivalent relation in $M_{n\times n}(\mathbb{R})$.

Similar matrices will play an important role. We will soon see that every matrix will be similar to relatively nice matrix so that their properties can be investigated, as well as making computations significantly easier.

3.3 Diagonalization

We now show a very nice kind of matrices, diagonal matrices that will come into play with linear maps. Our goal is to attempt to classify, by similarity, of every matrix into a diagonal one. We will soon see that this is not possible, and thus giving the last section of these notes meaning.

Definition 3.3.1: Diagonalizable Linear Maps

An linear map T is diagonalizable if there exists a basis the matrix representation of T is linear.

Proposition 3.3.2: Diagonalizable Matrices

A linear map T represented by A is diagonalizable if there exists an invertible P and a diagonal matrix D such that $P^{-1}AP = D$. In that case, P consists of eigenvectors of T and the diagonals of D are the eigenvalues of A.

Theorem 3.3.3

If the linear map $T : V \to V$ has n distinct eigenvalues where $\dim(V) = n$, then T is diagonalizable.

Although not stated in the theorem, this does not mean that linear maps without n distinct eigenvalues are not diagonalizable. However by taking the contrapositive, we see that not every linear map is diagonalizable because clearly, not every linear map has n distinct eigenvalues.

Theorem 3.3.4

Let $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then following are equivalent.

- T is diagonalizable
- V has a basis consisting of eigenvalues of T
- $V = E_{\lambda_1} + \cdots + E_{\lambda_m}$
- $\dim(V) = \dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_m})$

4 The Jordan Canonical Form

In the last section, we looked into what kinds of matrices can have "nice" looking matrix under some basis. We now provide a less "nice" looking form of a similar matrix. However, every matrix can be reduced to this relatively "nice" looking form, as long as the field is algebraically closed. This form is called the Jordan Normal Form.

4.1 The Minimal Polynomial

Theorem 4.1.1

Let $\mathbb{F}$ be a field. Let A be a $n \times n$ matrix over $\mathbb{F}$. Then there is some non-zero polynomial $p \in \mathbb{F}[x]$ of degree at most n^2 such that $p(A) = \mathbf{0}_n$.

Proof. Note that $\{I, A, \dots, A^{n^2}\}$ is linerarly dependent in the vector space of $n \times n$ matrices. Thus there exists constant $c_0, \dots, c_{n^2}$ that are not all zero such that

$$c_0 I + \cdots + c_{n^2} A^{n^2} = \mathbf{0}_n$$

Thus $p(x) = c_0 + c_1 x + \cdots + c_{n^2} x^{n^2}$ is our desired polynomial. $\square$

Theorem 4.1.2

Let $A_{n\times n}$ be a matrix over $\mathbb{F}$ representing the linear map $T : V \to V$. Then

- There is a unique monic non-zero polynomial $p(x)$ with minimal degree and coefficients in $\mathbb{F}$ such that $p(A) = \mathbf{0}_n$
- If $q(x)$ is any polynomial with $q(A) = \mathbf{0}_n$, then $p|q$

Proof. By the previous theorem, there exists a polynomial such that $p(A) = 0$. Divide the polynomial by c_{n^2} gives us the desired monic polynomial. Suppose that p_1, p_2 are distinct monic polynomials that are minimal such that $p_1(A) = 0$ and $p_2(A) = 0$, then $p = p_1 - p_2$ is a non zero polynomial with a smaller degree and $p(A) = 0$, contradicting the minimality of degree. Thus p is unique.

Let $p(x)$ be the minimal polynomial in the above proof. Let $q(A) = 0$. By division algorithm there exists some r with smaller degree than p such that $q = sp + r$. If r is non-zero, then $r(A) = q(A) - s(A)p(A) = 0$, contradiction of minimality, thus $r = 0$ and $p|q$. $\square$

Definition 4.1.3: The Minimal Polynomial

The unique monic non-zero polynomial $\mu_A(x)$ of minimal degree with $\mu_A(A) = \mathbf{0}_n$ is called the minimal polynomial of A.

Proposition 4.1.4

Similar matrices have the same minimal polynomial.

Proof. Similar matrices represent the same linear map, thus both have their minimal polynomial same as T, the linear map. $\square$

Proposition 4.1.5

Let D be a diagonal matrix with $\{d_1, \dots, d_r\}$ its unqiue diagonal entries, then

$$\mu_D(x) = (x - d_1) \cdots (x - d_r)$$

Proof. For any diagonal matrix,

$$p(D) = \begin{pmatrix} p(d_{11}) & 0 & \cdots & 0 \\ 0 & p(d_{22}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(d_{nn}) \end{pmatrix}$$

Thus $p(D) = 0$ if and only if $p(d_{kk}) = 0$ for $k \in \{1, \dots, n\}$. Thus the smallest-degree monic polynomial vanishing at these points is clearly the polynomial above. ☐

Corollary 4.1.6

Every diagonalizable matrix has its minimal polynomial a product of distinct linear factors.

Proof. Since diagonalizable matrix is similar to some diagonal matrix and they both have the same minimal polynomial, by the above proposition it is a product of distinct linear factors. ☐

We will later see that in fact, the above criterion is a neccessary and sufficient condition: A is diagonalizable if and only if the minimal polynomial is a product of distinct linear factors.

4.2 Cayley-Hamilton Theorem

Theorem 4.2.1: Cayley-Hamilton

Let $c_A(x) = \det(A - xI)$ be the characteristic polynomial of the $n \times n$ matrix A over a field $\mathbb{F}$, then $c_A(A) = 0$.

Proof. Firstly note that if $P(x) = \sum_{i=1}^n P_i x^i$ and $Q(x) = \sum_{j=1}^m Q_j x^j$ are polynomials with matrix coefficients where the matrix is $n \times n$, and $R(x) = \sum_{k=1}^{n+m} R_k x^k$ is the product of the two polynomials with $R_k = \sum_{i+j=k} P_i Q_j$, then if M is a $n \times n$ matrix that commmutes with all of Q_j, then we have

$$R(M) = P(M)Q(M)$$

This can be seen by expanding the sums out.

Now take $Q(x) = A - xI$ and $P(x) = \text{adj}(Q)$. Then we have
$P(x)Q(x) = \det(A - xI)I = c_A(x)I$ by property of the adjoint. And since A commutes with all coefficients of the polynomial of Q, we have

$$c_A(A)I = P(A)Q(A) = P(A) \cdot 0 = 0$$

Thus $c_A(A) = 0$. ☐

Corollary 4.2.2

For any $A_{n\times n}$ over $\mathbb{F}$, we have $\mu_A | c_A$, and $\deg(\mu_A) \leq n$.

Proof. This is clear since $c_A(A) = 0$ and μ_A is the minimal polynomial such that $\mu_A(A) = 0$ by division with remainder. Since $\deg(c_A) = n$, $\deg(\mu_A) \leq n$. □

This lemma may help with finding out the minimal polynomial.

Lemma 4.2.3

Let λ be an eigenvalue of A. Then $\mu_A(\lambda) = 0$.

Proof. Let v be an eigenvector of the eigenvalue λ of A. Trivially $\mu_A(A)v = 0$. But also since

$$A^n v = \lambda^n v$$

we have $0 = \mu_A(A)v = \mu_A(\lambda)v$. Since v is nonzero we must have $\mu_A(\lambda) = 0$. □

In general, to deduce the formula for the minimal polynomial, we follow three steps.

Step 1: Find out the eigenvalues of the matrix.
Step 2: List out the possibilities of the minimal degree. This is done using the fact that $\mu_A(\lambda) = 0$ and $\deg(\mu_A) \leq n$.
Step 3: Plug in the matrix to find out which polynomial has its root at A.

There is another method to find out the formula using the following lemma.

Lemma 4.2.4

Let $T : V \to V$ be a linear map. Let

$$V = W_1 \oplus \cdots \oplus W_k$$

be the direct sum of invariant subspaces, meaning $W_1, \dots, W_k$ are invariant subspaces of T. Let $\mu_i(x)$ be the minimal polynomial of $T|_{W_i}$. Then

$$\mu_T(x) = \operatorname{lcm}(\mu_1, \dots, \mu_k)$$

Using this, we derive a better algorithm to find the minimal polynomial:

Step 1: Take $v \neq 0$ an eigenvector and set $W = \operatorname{span}\{v, T(v), T^2(v), \dots\}$. Then W is invariant under T. Let d be the minimal positive integer such that $v, T(v), \dots, T^d(v)$ are linearly dependent. Then $v, T(v), \dots, T^{d-1}(v)$ are linearly independent. Then we know that $\mu_T(x)$ has degree larger than d since else $\mu_T(x)v$ will never be 0. Then there is a nontrivial linear dependency relation of the form

$$T^d(v) + c_{d-1}T^{d-1}(v) + \cdots + c_1 T(v) + c_0 v = 0$$

Step 2: Consider the polynomial

$$x^d + c_{d-1}x^{d-1} + \cdots + c_1 x + c_0$$

Then this is precisely the minimal polynomial.

4.3 Generalized Eigenspace

Definition 4.3.1: Generalized Eigenvector

Let $T : V \to V$. Fix $k \in \mathbb{N}$. A non zero vector v such that

$$(T - \lambda I)^k v = 0$$

is called a generalized eigenvector of T with respect to the eigenvalue λ. Also we define

$$N_k(T,\lambda)=\{v\in V|(T-\lambda I)^k v=0\}=\ker((T-\lambda I)^k)$$

to be the generalized eigenspace of index k of T with respect to λ. The set of all generalized eigenvector regardless of the index, is defined to be

$$G(T,\lambda)=\{v\in V|(T-\lambda I)^k v=0 \text{ for some } k\in\mathbb{N}\}=\bigcup_{k=1}^{\infty}N_k(T,\lambda)$$

Proposition 4.3.2

The dimensions of corresponding generalized eigenspaces of similar matrices are the same.

Proof. This is true since generalized eigenspaces are defined without explicitly defining a basis for the linear map. Thus similar matrices that induce the same linear map will have the same dimensions for generalized eigenspaces. □

Proposition 4.3.3

Let $T:V\to V$ be a linear map with eigenvalue λ. Then

$$N_1(T,\lambda)\subseteq N_2(T,\lambda)\subseteq N_3(T,\lambda)\subseteq\dots$$

Proof. Trivially if $v\in\ker(A-\lambda I)^i$. This means that $(A-\lambda I)^i v=0$ and $(A-\lambda I)^{i+1}v=0$. Thus $N_i(T,\lambda)\subseteq N_{i+1}(T,\lambda)$ for any i and we are done. □

Proposition 4.3.4

Let λ be an eigenvalue of $T:V\to V$. There exists some $n\in N$ such that

$$N_n(T-\lambda I)=N_{n+1}(T-\lambda I)=\dots$$

Denote $d(\lambda)$ the smallest of such n.

Proof. $d(\lambda)\leq\dim(V)$. □

Proposition 4.3.5

Let λ be an eigenvalue of $T:V\to V$. Then

$$G(T,\lambda)=N_{\dim(V)}(T,\lambda)$$

Proposition 4.3.6

Let $T:V\to V$ with eigenvalues $\lambda_1,\dots,\lambda_m$. Let $v_i\in G(T,\lambda_i)$ for $i\in\{1,\dots,m\}$. Then $v_1,\dots,v_m$ are linearly independent.

Theorem 4.3.7

Let V be a vector space of an algebraically closed field F. Let $T:V\to V$. Let $\lambda_1,\dots,\lambda_m$ be distinct eigenvalues of T. Then

- $V = G(T, \lambda_1) \oplus \cdots \oplus G(T, \lambda_m)$
- $G(T, \lambda_j)$ is invariant under T.

4.4 Jordan Canonical Form

Definition 4.4.1: Jordan Chain

A Jordan Chain of length k is a sequence of nonzero vectors $v_1, \dots, v_k$ such that

$$\begin{aligned} Av_1 &= \lambda v_1 \\ Av_2 &= \lambda v_2 + v_1 \\ &\vdots \\ Av_k &= \lambda v_k + v_{k-1} \end{aligned}$$

for some eigenvalue λ of A.

Corollary 4.4.2

Let $v_1, \dots, v_k$ be a Jordan Chain of λ. Then $v_i \in N_i(A, \lambda)$ for $i \in \{1, \dots, k\}$ and $(T - \lambda I)(v_i) = v_{i-1}$ except for $i = 1$.

Proof. The result is immediate from substitution in the Jordan Chains. □

Proposition 4.4.3

The vectors in a Jordan chain are linearly independent.

Proposition 4.4.4

The subspace spanned by a Jordan Chain is invariant under its linear map.

Proof. Note that we just have to find out where $v_1, \dots, v_k$ are mapped to since they are a basis of our subspace. But $T(v_i) = \lambda v_i + v_{i-1}$ for $i \in \{2, \dots, k\}$ which is a linear combination of our basis, we must have that the subspace is invariant. □

Definition 4.4.5: Jordan Block of Degree k

Define the Jordan block of degree k to be the $k \times k$ matrix

$$\gamma_{ij} = \begin{cases} \lambda & \text{if } j = i \\ 1 & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}$$

This means the diagonal of the matrix is λ and the super diagonal is 1. It is denoted as $J_{\lambda,k}$

Corollary 4.4.6

The matrix of T with respect to the basis $v_1, \dots, v_n$ is a Jordan Block if and only if $v_1, \dots, v_n$ is a Jordan Chain.

Proof. Let $v_1, \ldots, v_k$ be a Jordan Chain. Our matrix should be in the form

$$\begin{pmatrix} T(v_1) & T(v_2) & \cdots & T(v_k) \end{pmatrix}$$

Calculating each column gives

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

For the other side, it is easy to simply compute $v_1, \ldots, v_k$ out with the matrix. □

Definition 4.4.7: Jordan Basis

A Jordan basis is a basis consisting of one or more Jordan chains strung together.

Lemma 4.4.8

A Jordan Basis is indeed a basis.

Proof. We naturally assume the string of Jordan Chains consists of n vectors in total, corresponding to $\dim(V) = n$. Thus we just have to show linear independence. But this is also trivial. We have shown that vectors in the same Jordan Chain are independent, and vectors in different $G(T, \lambda)$ are proven to be linearly independent. □

Definition 4.4.9: Direct Sum

Let $A \in F^{n\times n}$ and $B \in F^{m\times m}$. Define the direct sum to be

$$A \oplus B = \begin{pmatrix} A & 0_{n\times m} \\ 0_{m\times n} & B \end{pmatrix}$$

Lemma 4.4.10

Let B, C be square matrices.

$$(B \oplus C)^n = B^n \oplus C^n$$

Corollary 4.4.11

The matrix of T with respect to a Jordan Basis is the direct sum

$$J_{\lambda_1,k_1} \oplus \cdots \oplus J_{\lambda_s,k_s}$$

of Jordan Blocks.

Theorem 4.4.12

Let A be a matrix over an algebraically closed field. Then there exists a Jordan basis for A. Moreover, A is similar to some J which is a direct sum of Jordan Blocks.

Proof. We will construct the basis with 3 methods. They each contribute to part of the Jordan Basis. Firstly, we will obtain a basis of a subspace. We induct on n. The case of $n = 1$ is trivial.

Now suppose $T : V \to V$ is a linear map with $\dim(V) = n$. We want to find a restriction of T that is an automorphism to apply the induction hypothesis. Fix λ to be one of the eigenvalues of T. This will be used throughout the entire proof. This λ is possible because the ground field is algebraically closed. Now I claim that $U = \text{im}(T - \lambda I)$ is invariant under T. If this is true, then $T|_U : U \to U$ can be used to apply induction hypothesis. So all we have to show is that U is T-invariant and that $\dim(U) < \dim(V)$. The second item must be true by the rank nullity theorem. There must be an eigenvector in $\ker(T - \lambda I)$. Thus $\ker(T - \lambda I) \geq 1$ which implies that $\dim(U) < n$ by the rank nullity theorem. Now we prove that U is invariant. Let $u \in U$, I show that $T(u) \in U$. If $u \in U$, then $u = (T - \lambda I)(v)$ for some $v \in V$, hence

$$T(u) = T(T - \lambda I)(v) = (T - \lambda I)(T(v)) \in \text{im}(T - \lambda I) = U$$

Thus we have proven that $T|_U : U \to U$. Apply induction hypothesis here to obtain a Jordan Basis for T_U. Call that Jordan Basis $e_1, \dots, e_m$.

We now construct our second set of vectors. Recall that a Jordan Basis is a string of l Jordan Chains. Let $v_1, \dots, v_k$ denote one of the Jordan Chains. We can extend this Jordan Chain one more by setting $(T - \lambda I)^{k+1}(v_{k+1}) = 0$. Do the same thing for everey Jordan Chain, and relabel them to $w_1, \dots, w_l$. As a side note, the two set of vectors we have now still form a Jordan Basis because we simply extended every Jordan Chain one more.

For the final set of vectors, observe that the first vector of each of the l Jordan Chains are eigenvectors of $T|_U$ with its eigenvalue being λ. This is because by definition of Jordan Chains, $T|_U(v_1) = \lambda v_1$. Also note that those l vectors are linearly independent. Thus the first vectors of each of the l Jordan Chains span an l dimensional subspace of the eigenspace of λ. Recall that the eigenspace of λ has dimension $\dim(V) - \dim(U) = \dim(\ker(T - \lambda I))$. To minimize notation let $m = \dim(U)$. Thus by extension theorem we can extend the basis of the l dimensional subspace to $\ker(T - \lambda I)$. Call the extension vectors $w_{l+1}, \dots, w_{n-m}$. As a side note, these $n - m - l$ vectors each Jordan Chains of length 1. Thus we have complete our last set of vectors.

We have n vectors

$$e_1, \dots, e_m, w_1, \dots, w_l, w_{l+1}, \dots, w_{n-m}$$

Thus we only need to prove that they are linearly independent. Let $x = \sum_{k=1}^{m} \beta_m e_m$. Let

$$\sum_{i=1}^{n-m} \alpha_i w_i + x = 0$$

Applying $T - \lambda I$ on both sides give

$$\sum_{i=1}^{l} \alpha_i (T - \lambda I) w_i + (T - \lambda I)(x) = 0$$

Since $w_{l+1}, \dots, w_{n-m}$ is in $\ker(T - \lambda I)$, they become 0. Now recall that our construction of $w_1, \dots, w_l$ is made by extending our Jordan Chains. So applying $(T - \lambda I)$ moves down our Jordan Chain. This means that $(T - \lambda I)x$ no longer contains the last term of each Jordan Chain and are linear combinations of $\{e_1, \dots, e_m\} \setminus \{\text{Last Term of each Jordan Chain}\}$, while all of the $(T - \lambda I)(w_1), \dots, (T - \lambda I)(w_l)$ are all last members of ecah Jordan Chain. From the fact that $e_1, \dots, e_m$ are a basis, we have $\alpha_1 = \dots = \alpha_l = 0$.

Our sum now becomes $(T - \lambda I)x = 0$. Which means that $x \in \ker(T - \lambda I)$. Now our original sum becomes

$$\sum_{i=l+1}^{n-m} \alpha_i w_i + x = 0$$

By construction, $w_{l+1}, \dots, w_{n-m}$ extends a basis of the eigenspace of $T|_U$ for λ, thus $\alpha_{l+1} = \dots = \alpha_{n-m} = 0$. Also since $e_1, \dots, e_m$ is a basis of U, we have $\beta_1 = \dots = \beta_m = 0$. Finally by the above corollary, in a Jordan Basis, the matrix of T is a direct sum of Jordan Blocks. $\square$

Lemma 4.4.13

If A, B are similar, then they have the same JCF up to reodering of the Jordan Blocks by direct sum.

Theorem 4.4.14

Let λ be an eigenvalue of a matrix A. Let J be the Jordan Canonical Form of A. Then

- The number of Jordan Blocks of J with eigenvalue λ is equal to $\dim(\ker(A - \lambda I))$
- Let $k > 0$. Then number of Jordan Blocks of J with eigenvalue λ of degree at least i is equal to $\dim(N_i(A, \lambda)) - \dim(N_{i-1}(A, \lambda))$

Proof. Since similar matrices have the same dimensions for their generalized eigenspaces corresponding to their eigenvalue, WLOG take $A = J = J_{\lambda_1,k_1} \oplus \cdots \oplus J_{\lambda_s,k_s}$. However, note that the dimension of $N_i(A \oplus B, \lambda)$ is equal to $\dim(N_i(A, \lambda)) + \dim(N_i(B, \lambda))$. So we just have to prove the theorem for a single Jordan Block.

Since $(J_{\lambda,k} - \lambda I)^i$ has a single diagonal line of ones i places above the diagonal for $i < k$, and is 0 for $i \geq k$, the dimension of its kernel is i for $0 \leq i \leq k$ and k for $i \geq k$. $\square$

Corollary 4.4.15

The JCF of a matrix is unique up to a reordering of the Jordan Blocks.

Proof. The above theorem says that the number of Jordan Blocks associated with λ is determined by the nullity of A, and the size of every Jordan Block is determined by the dimension of the generalized eigenspaces. $\square$

4.5 Results of the Jordan Normal Theorem

Lemma 4.5.1

Let $M = A \oplus B$. Then

$$c_M(x) = c_A(x)c_B(x)$$

and

$$\mu_M(x) = \gcd(\mu_A(x), \mu_B(x))$$

Proposition 4.5.2

Let A have JCF J. Let λ be an eigenvalue of A. Consider the Jordan Blocks in J related to λ. The string of Jordan Chains of these Jordan Blocks form a basis for $G(A, \lambda)$.

Theorem 4.5.3

Let $T : V \to V$ and $\lambda_1, \dots, \lambda_m$ be the set of eigenvalues of T. Then the characteristic polynomial of T is

$$c_A(x) = (-1)^n \prod_{k=1}^{m} (x - \lambda_k)^{a_k}$$

where a_k is the sum of the degrees of the Jordan Blocks of T of eigenvalue λ_k

Theorem 4.5.4

Let $T : V \to V$ and $\lambda_1, \dots, \lambda_m$ be the set of eigenvalues of T. Then the minimal polynomial of T is

$$\mu_A(x) = \prod_{k=1}^{m} (x - \lambda_k)^{b_k}$$

where b_k is the largest among the degrees of the Jordan Blocks of T of eigenvalue of λ_k. Also, we have $d(\lambda_k) = b_k$

Theorem 4.5.5

Let $T : V \to V$ and $\lambda_1, \dots, \lambda_m$ be the set of eigenvalues of T. Then T is diagonalizable if and only if $\mu_A(x)$ has no repeated factors.

Theorem 4.5.6

Let $A, B \in M_{n\times n}(\mathbb{C})$. Then A and B are similar if and only if the following two are true:

- A and B have the same set of eigenvalues
- $\dim(\ker(A - \lambda I)^i) = \dim(\ker(B - \lambda I)^i)$ for all i and eigenvalues λ

To finish this section, we show the process of determining the Jordan Canonical form of a matrix. The steps are usually as follows:

Step 1: Find out the nullity of $A - \lambda I$ as this gives us the number of Jordan Blocks with eigenvalue λ.
Step 2: To find out the number of Jordan blocks with eigenvalue λ and size at least i, we calculate $\dim(\ker(A - \lambda I_n)^i) - \dim(\ker(A - \lambda I)^{i-1})$.

To find out the change of basis matrix, meaning the Jordan Chains, we do the following step:

Step 3: Find out the last one in the chain v_k by solving $(A - \lambda I)^k v_k = 0$ while restricting v_k such that $(A - \lambda I)^{k-1} v_k \neq 0$, and then proceed to find out v_{k-1} by $(A - \lambda I)^{k-1} v_k = v_{k-1}$ and vice versa.

There are also extra information that we can use to determine the JCF:
The degree of $(x - \lambda)$ in μ_A indicates the maximum size of Jordan Blocks with eigenvalue λ.
The degree of $(x - \lambda)$ in c_A indicates the total size of used in the JCF of all Jordan Blocks with eigenvalue λ.

We give an example of finding the JCF of a matrix.

Example 4.5.7

Find the Jordan Canonical Form of

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

Proof. We begin by finding out the eigenvalues of A. We have that $c_A(x) = \det(A - xI) = (1 - x)^2(2 - x)$. This means that the eigenvalues are 1 and 2. Now we begin with step 1.

For eigenvalue 1, we have that row reduced form of $A - I$ is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus the nullity of $A - I$ is 1. This means that the number of Jordan blocks with eigenvalue 1 is 1. Using information from c_A, we know that the total size used for the eigenvalue 1 is 2. This means that there is exactly one Jordan block of size 2 in the JCF of A.

This leaves the fact that the remaining Jordan block of size 1 being the eigenvalue 2.

With this, we complete the JCF of A with

$$J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

To compute the basis, or the change of basis matrix P, we use step 3. Since A has one Jordan block for eigenvalue 1, we need to find one string of Jordan chain. This chain needs to have length 2 since the size of the Jordan block is 2. (If there are multiple of Jordan blocks of the same eigenvalues, the end vector of the Jordan chains needs to be linearly indendent). We begin by finding the ending of the chain, v_2 by using the fact that $(A - I)^2 v_2 = 0$ and $(A - I)v_2 = v_1 \neq 0$. We have that

$$(A - I)^2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

We choose that $v_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$. Now we have

$$v_1 = (A - I)v_2 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

Finally, we choose $v_3 \in \ker(A - 2I)$. But row reducing $A - 2I$ gives

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

We can choose $v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. This means that

$$P = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

$\square$

4.6 Functions of Matrices

We start of the last section with a formula for the nth power of a Jordan Block.

Lemma 4.6.1

Let $J_{\lambda,k}$ be a Jordan Block. Then

$$J_{\lambda,k}^n = \begin{pmatrix} \lambda^n & n\lambda^{n-1} & \cdots & \binom{n}{k-2}\lambda^{n-k+2} & \binom{n}{k-1}\lambda^{n-k+1} \\ 0 & \lambda^n & \cdots & \binom{n}{k-3}\lambda^{n-k+3} & \binom{n}{k-2}\lambda^{n-k+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda^n & n\lambda^{n-1} \\ 0 & 0 & \cdots & 0 & \lambda^n \end{pmatrix}$$

When used together with the fact that powers of matrices can be distributed through direct sums, we obtain the formula for finding powers of matrices. Namely if A has Jordan canonical form J and $A = PJP^{-1}$, then $A^n = PJ^nP^{-1}$.

We now define functions of matrices in terms of this decomposition.

Definition 4.6.2: Functions of Matrices

Let f be a function over $\mathbb{C}$. For every matrix $A \in M_{n\times n}(\mathbb{C})$, define $f(A)$ by

$$f(A) = Pf(J)P^{-1}$$

where $f(J) = f(J_{\lambda_1,k_1}) \oplus \cdots \oplus f(J_{\lambda_t,k_t})$ is the direct sum of Jordan blocks. And finally, for each Jordan block $J_{\lambda,k}$, define $f(J_{\lambda,k})$ by

$$f(J_{\lambda,k}) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) & \frac{1}{(k-1)!}f^{(k-1)}(\lambda) \\ 0 & f(\lambda) & \cdots & \frac{1}{(k-3)!}f^{(k-3)}(\lambda) & \frac{1}{(k-2)!}f^{(k-2)}(\lambda) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & f(\lambda) & f'(\lambda) \\ 0 & 0 & \cdots & 0 & f(\lambda) \end{pmatrix}$$

Note that taking $f(x) = x^n$, this coincides with the power of a Jordan block, which is presumably what motivated the this definition of matrix functions.

5 Linear Forms and Quadratic Forms

5.1 Linear Forms

Definition 5.1.1: Linear Forms

A linear form on V is a linear map from V to $\mathbb{F}$.

Proposition 5.1.2: Dual Space

The set of all linear forms on V forms a vector space called the dual space V'.

Proof. Simply a check on the axioms of vector space. □

Lemma 5.1.3

Let V be a finite dimensional vector space. Then V' is also finite dimensional and $\dim(V') = \dim(V)$.

Definition 5.1.4: Dual Basis

Let $v_1, \dots, v_n$ be a basis of V, then the dual basis of $v_1, \dots, v_n$ is the list $\phi_1 \dots, \phi_n$ of elements of V', where ϕ_k is a linear functional such that

$$\phi_k(v_i) = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i \end{cases}$$

Proposition 5.1.5

The dual basis of a basis of V is a basis of V'

Definition 5.1.6: Dual Map

Let $T \in \mathcal{L}(V, W)$. The dual map of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\phi) = \phi \circ T$ for $\phi \in W'$.

Proposition 5.1.7

Let $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$.

- $(S + T)' = S' + T'$
- $(\lambda T)' = \lambda T'$
- $(ST)' = T'S'$.

5.2 Quadratic Forms

Definition 5.2.1: Quadratic Forms

A quadratic form in n variables $x_1, \dots, x_n$ over a field K is a polynomial

$$q(x_1, \dots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Proposition 5.2.2

Every quadratic form $q(x_1,\dots,x_n)=\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j$ can be represented by a matrix multiplication, namely

$$q(x_1,\dots,x_n)=\begin{pmatrix}x_1 & \cdots & x_n\end{pmatrix}\begin{pmatrix}a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n}\\ \frac{1}{2}a_{21} & a_{22} & \cdots & \frac{1}{2}a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{2}a_{n1} & \frac{1}{2}a_{n2} & \cdots & a_{nn}\end{pmatrix}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$$

In particular, this matrix is symmetric with $a_{ij}=a_{ji}$ for $i,j\in\{1,\dots,n\}$.

Proof. Multiplying out the entries of the matrix multiplication gives the original quadratic form. □

Proposition 5.2.3

A change of basis via the change of basis matrix P also changes the symmetric matrix of the quadratic form by P^TAP

Definition 5.2.4: Congruent Matrices

Two matrices A,B are said to be congruent if there exists some invertible matrix P such that $B=P^TAP$

Beware that congruences does not apply to only symmetric matrices. We will see more of it in action in bilinear forms.

Proposition 5.2.5

Two symmetric matrices are congruent if and only if they represent the same quadratic form with respect to different bases.

Theorem 5.2.6

Let $q(x_1,\dots,x_n)$ be a quadratic form in n variables over a field K whose characteristic is not 2. Then there exists a basis such that $q(y_1,\dots,y_n)=c_1y_1^2+\cdots+c_ny_n^2$ for some $c_1,\dots,c_n\in K$.

Proof. There is a shorter proof for this theorem, but for the sake of the construction of $c_1,\dots,c_n$, we will prove the theorem constructively. Suppose that q is represented by the symmetric matrix $A=(a_{ij})_{n\times n}$ with respect to the basis $b_1,\dots,b_n$. There are three steps in the construction. I use $b_1,\dots,b_n$ to indicate the old basis and $b'_1,\dots,b'_n$ to indicate the basis after the step.

Step 1: Arrange such that $q(b_1)\neq 0$. There are four cases here.

- If $a_{11}\neq 0$, then we are done.
- If $a_{11}=0$ but $a_{kk}\neq 0$ for some $1<k\leq n$. Then just set $b'_1=b_k$ and $b'_k=b_1$. At the same time, the matrix for the quadratic form is changed by swapping rows r_1 and r_k, and then swapping the columns c_1 and c_k
- If $a_{kk}=0$ for all $k\in\{1,\dots,n\}$, but there are some i,j such that $a_{ij}\neq 0$, then set $b'_i=b_i+b_j$ since $q(b_i+b_j)=2a_{ij}\neq 0$ and so we reduced this case to the previous two

cases.The matrix then becomes

$$\begin{pmatrix} 2a_{1k} & a_{12}+a_{k2} & \dots & a_{1k} & \cdots & a_{1n}+a_{kn} \\ a_{12}+a_{k2} & 0 & \cdots & a_{2k} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1k} & a_{k2} & \cdots & 0 & \cdots & a_{kn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1n}+a_{kn} & a_{n2} & \cdots & a_{nk} & \cdots & 0 \end{pmatrix}$$

- If $a_{ij}=0$ for all $i,j\in\{1,\dots,n\}$ then it is the zero function.

In this step the change of basis matrix is just the elementary matrices.

Step 2: Now we modify $b_2,\dots,b_n$ to make them orthogonal to b_1. Now set $b_k'=b_k-\frac{a_{1k}}{a_{11}}b_1$. This way, the matrix entry a_{1k} becomes zero. Now the change of basis matrix becomes

$$P=\begin{pmatrix} 1 & -\frac{a_{12}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

After this step all the change of basis matrix should be compiled and calculated so that the new matrix for the quadratic form can be formed.

Step 3: Since the matrix for the quadratic form is now

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & ? & \cdots & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & ? & \cdots & ? \end{pmatrix}$$

We can induct on n by repeating the process of step 1 with the entry a_{22} until we reach a_{nn}. □

This main theorem of quadratic forms shows that every quadratic form is congruent to a diagonal matrix. To illustrate the process of reduction, we look an example.

Example 5.2.7

Find a nice basis for the quadratic form $q\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right)=xy+3yz-5xz$.

Proof. Using the formula, we construct the matrix of q as

$$A=\begin{pmatrix} 0 & \frac{1}{2} & -\frac{5}{2} \\ \frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{5}{2} & \frac{3}{2} & 0 \end{pmatrix}$$

We start the first change of basis. Notice that $a_{11}=a_{22}=a_{33}=0$ in the matrix while a_{12} is not. Denote the standard basis by b_1,b_2,b_3. We perform the first basis change by $\{b_1'=b_1+b_2,b_2'=b_2,b_3'=b_3\}$. This means that the change of basis matrix from new to old is

$$P'=\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

To change A into the new matrix A', we simply replace r_1 by $r_1 + r_2$ and replace c_1 by $c_1 + c_2$. This gives

$$A' = \begin{pmatrix} 1 & \frac{1}{2} & -1 \\ \frac{1}{2} & 0 & \frac{3}{2} \\ -1 & \frac{3}{2} & 0 \end{pmatrix}$$

We keep track of changing the basis for clarity. We now have $A' = (P')^T A P'$.

The next step is to set the new basis to $b_2'' = b_2' - \frac{1}{2}b_1'$ and $b_3'' = b_3' + b_1'$. This means that the new basis is now $\{b_1 + b_2, \frac{1}{2}(b_2 - b_1), b_1 + b_2 + b_3\}$. The change of basis from this basis back to the old one is now

$$P'' = \begin{pmatrix} 1 & -\frac{1}{2} & 1 \\ 1 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Now the new matrix A'' is formed by replacing r_2 with r_2 by $r_2 - \frac{1}{2}r_1$ and r_3 with $r_3 + r_1$. Noticing that A'' must be symmetric, we need to take the new elements and replace the remanining lower traingualr elements so that A'' maintains symmetric. Also observe that the replacement of rows is exactly one changes into the new basis from the previous basis, where $b_2'' = b_2' - \frac{1}{2}b_1'$ etc. Now we have

$$A'' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{4} & 2 \\ 0 & 2 & -1 \end{pmatrix}$$

We now have $A'' = (P'')^T A P''$.

Now we perform the next change of basis. We set $b_3''' = b_3'' - \frac{2}{-\frac{1}{4}}b_2'' = b_3'' + 8b_2''$. Now the new basis is $\{b_1 + b_2, \frac{1}{2}(b_2 - b_1), -3b_1 + 5b_2 + b_3\}$. The change of basis matrix is now

$$P''' = \begin{pmatrix} 1 & -\frac{1}{2} & -3 \\ 1 & \frac{1}{2} & 5 \\ 0 & 0 & 1 \end{pmatrix}$$

Similar to the above, we replace r_3 with the same transformation and adjust A''' so that it remains symmetric. Thus now we have

$$A''' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{4} & 0 \\ 0 & 0 & 15 \end{pmatrix}$$

This means that we are done with $A''' = (P''')^T A P'''$. $\square$

In general, this result of diagonalization is different from that of similar matrices. One should not be confusing reduction of congruent matrices into diagonal matrices and reduction of similar matrices into JCF as well as reduction of diagonalizable matrices into diagonal matrices.

5.3 Bilinear Forms

Definition 5.3.1: Bilinear Maps

Let V, W be vector spaces over a field $\mathbb{F}$. A bilinear map on V and W is a map $\tau : V \times W \to \mathbb{F}$ such that

- $\tau(a_1 + a_2v_2, w) = a_1\tau(v_1, w) + a_2\tau(v_2, w)$
- $\tau(v, b_1w_1 + b_2w_2) = b_1\tau(v, w_1) + b_2\tau(v, w_2)$

Theorem 5.3.2

Let V and W has basis $e_1, \dots e_n$ and $f_1, \dots, f_m$ respectively. Let $\tau(v, w)$ be a bilinear map. Then

$$\tau(v,w) = v^T \begin{pmatrix} \tau(e_1, f_1) & \cdots & \tau(e_1, f_m) \\ \vdots & \ddots & \vdots \\ \tau(e_n, f_1) & \cdots & \tau(e_n, f_m) \end{pmatrix} w$$

Proof. Simple to see by expanding the matrix multiplication. □

Proposition 5.3.3

Let A be the matrix of the bilinear map $\tau V \times W \to \mathbb{F}$ with respect to the basis $e_1, \dots, e_n$ and $f_1, \dots, f_m$ of V and W. Let B be the matrix with respect to the basis of $e'_1, \dots, e'_n$ and $f'_1, \dots, f'_m$ of V and W. Let P and Q be the change of basis matrix where $Pv' = v$ and $Qw' = w$ for $v \in V$ and $w \in W$. Then

$$B = P^T AQ$$

Definition 5.3.4: Bilinear Forms

A bilinear form is a bilinear map that maps $V \times V$ to $\mathbb{F}$.

Definition 5.3.5: Congruent Matrices

Two matrices are congruent if there exists an invertible matrix P such that $B = P^T AP$.

Note that this definition of congruence in matrices coincides with the definition given with symmetric matrices.

Definition 5.3.6: Types of Bilinear Forms

A bilinear form is said to be

- symmetric if $\tau(v, w) = \tau(w, v)$
- antisymmetric if $\tau(v, w) = -\tau(w, v)$
- positive definite if $\tau(v, v) > 0$ for all $v \in V$.

Proposition 5.3.7

Let $q : V \to K$ be a function. Then the following are equivalent.

- q is a quadratic form
- $q(cv) = c^2 v$ for $c \in K$ and $v \in V$ and $\tau(v, w) = \frac{1}{2}(q(v + w) - q(v) - q(w))$ is a bilinear form on V
- $q(v) = \tau(v, v)$ is a symmetric bilinear form on V

Proof. Let $q : V \to K$ be a function.

- (1) $\implies$ (2): Since every term in a quadratic form is quadratic, $q(cv) = c^2 q(v)$ is natural. The fact that $\tau(v, w)$ is bilinear is also easy to check.

- (2) $\implies$ (3): From (2) we know that $\tau(v, v) = q(v)$ by substituting v in the position of w and thus it clearly is a bilinear form. The position of w and v are also interchangable and thus it is symmetric.
- (3) $\implies$ (1): If $\tau(v, v)$ is a symmetric bilinear form then the matrix of τ is symmetric since $a_{ij} = \tau(e_i, f_j) = \tau(f_j, e_i) = a_{ji}$. Thus $q(v) = \tau(v, v)$ defines a quadratic form.

$\square$

5.4 Sesquilinear Forms

6 Inner Product Spaces

6.1 Norms

Throught this section, $\mathbb{F}$ means either $\mathbb{R}$ or $\mathbb{C}$. In general normed vector spaces only perform well in these two fields.

Definition 6.1.1: Norm

Let V be a vector space. A norm on V is a function $\|\cdot\| : V \to \mathbb{F}$ such that

- $\|x\| \geq 0$ for all $x \in V$ with equality if and only if $x = 0$
- $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in V$ and $\lambda \in \mathbb{F}$
- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$

Definition 6.1.2: Normed Vector Space

A normed vector space is a pair $(V, \|\cdot\|)$ where V is a vector space and $\|\cdot\|$ is a norm on V.

Proposition 6.1.3

Every normed vector space is a metric space.

Proof. Can easily be seen that setting $d(x, y) = \|x - y\|$ allows the norm to become a metric. □

Definition 6.1.4: Convex Set

Let V be a vector space. A subset K of V is convex if $x, y \in K$ implies

$$\lambda x + (1 - \lambda)y \in K$$

for $0 \leq \lambda \leq 1$.

Lemma 6.1.5

For every normed vector space, the unit ball $B_1(0) = \{v \in V | \|v\| \leq 1\}$ is convex.

Proposition 6.1.6

Let $N : V \to \mathbb{R}^+$ be a function that satisfies the first two requirements of a norm. If N also satisfies the fact that $\{x \in V | N(x) \leq 1\}$ is convex, then N is a norm.

Definition 6.1.7: Isometries

If (X, d_1) and (Y, d_2) are metric spaces, then a distancing preserving map between X and Y is a map

$$f : X \to Y$$

such that for any $P, Q \in X$, we have

$$d(f(P), f(Q)) = d(P, Q)$$

A bijective distancing preserving map is called an isometry.

6.2 Inner Products

Inner products are only properly defined for vector spaces over $\mathbb{R}$ and $\mathbb{C}$. From this point onwards we will limit our discussions with $V = \mathbb{R}^n$ or $\mathbb{C}^n$ and $K = \mathbb{R}$ or $\mathbb{C}$.

Definition 6.2.1: Inner Products

An inner product on V is a function that takes each ordered pair (x, y) of a vector space V to a number $\langle x, y\rangle \in K$ such that

- $\langle x, x\rangle \geq 0$ for all $x \in V$ with equality if and only if $x = 0$.
- $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$ for all x for all $x, y, z \in V$
- $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $x, y \in V$ and $\lambda \in K$
- $\langle x, y\rangle = \overline{\langle y, x\rangle}$ for all $x, y \in V$

In this case V is called an inner product space.

Proposition 6.2.2

Let $u, v, w \in V$. Let $\lambda \in K$. Then the following are true.

- $\langle 0, u\rangle = \langle u, 0\rangle = 0$
- $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$
- $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$

Proof. Let $\langle\cdot,\cdot\rangle$ be an inner product over V.

- $\langle 0, u\rangle = \langle 0, u\rangle + \langle 0, u\rangle$ thus $\langle 0, u\rangle = 0$. $\langle u, 0\rangle = 0$ can be proven using the below property.
- Let $u, v, w \in V$. Then

$$\begin{aligned}
\langle u, v + w\rangle &= \overline{\overline{\langle u, v + w\rangle}} \\
&= \overline{\langle v + w, u\rangle} \\
&= \overline{\langle v, u\rangle + \langle w, u\rangle} \\
&= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle} \\
&= \langle u, v\rangle + \langle u, w\rangle
\end{aligned}$$

- Applying the same technique as above gives the desired result.

□

Proposition 6.2.3

Every inner product induces a norm.

Proof. Define the norm to be $\|x\| = \sqrt{\langle x, x\rangle}$. □

Proposition 6.2.4: Cauchy-Schwartz Inquality

For all $x, y \in V$,

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

with equality if and only if $y = \lambda x$ for some $\lambda \in \mathbb{F}$.

Proof. Let $z = x - \frac{|\langle x,y\rangle|}{\|y\|^2}y$. We have $\|z\| \geq 0$.

$$\begin{aligned}
\|z\|^2 &= \left\langle x - \frac{|\langle x,y\rangle|}{\|y\|^2}y, x - \frac{|\langle x,y\rangle|}{\|y\|^2}y \right\rangle \\
&= \langle x, x\rangle - 2\left\langle x, \frac{|\langle x,y\rangle|}{\|y\|^2}y \right\rangle + \left\langle \frac{|\langle x,y\rangle|}{\|y\|^2}y, \frac{|\langle x,y\rangle|}{\|y\|^2}y \right\rangle \\
&= \langle x, x\rangle - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^4}\langle y, y\rangle \\
&= \langle x, x\rangle - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^2} \\
&= \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}
\end{aligned}$$

Thus we have

$$\begin{aligned}
\|x\|^2 &\geq \frac{|\langle x,y\rangle|^2}{\|y\|^2} \\
\|x\| &\geq \frac{|\langle x,y\rangle|}{\|y\|} && \text{(Since they are all positive)} \\
\|x\| \cdot \|y\| &\geq |\langle x,y\rangle|
\end{aligned}$$

Note that we have equality if and only if $\|z\| = 0$, meaning $x = \frac{|\langle x,y\rangle|}{\|y\|^2}y$. We are done by taking $\lambda = \frac{\|y\|^2}{|\langle x,y\rangle|}$. $\square$

7 Orthogonality

7.1 Orthogonal Vectors

Definition 7.1.1: Orthogonal Vectors

Two vectors $u, v \in V$ an inner product space are called orthogonal if $\langle u, v\rangle = 0$.

Corollary 7.1.2

Let V be an inner product space. Let $u, v \in V$. Then the following are true.

- 0 is orthogonal to every vector in V
- 0 is the only vector in V that is orthogonal to itself.

Proof. Easy check involving properties of the inner product. □

Theorem 7.1.3: Pythagorean Theorem

Suppose that $u, v \in V$ an inner product space and u, v are orthogonal. Then

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

Proof. The norm here is induced by the inner product and thus $\|x\| = \sqrt{\langle x, x\rangle}$. We have that

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \langle u, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \|v\|^2\end{aligned}$$

□

Theorem 7.1.4: Orthogonal Decomposition

Let $u, v \in V$ and $v \neq 0$. Set $c = \frac{\langle u,v\rangle}{\|v\|^2}$ and $w = u - \frac{\langle u,v\rangle}{\|v\|^2}v$. Then $\langle w, v\rangle = 0$ and $u = cv + w$.

Proof. The fact that $u = cv + w$ is natural so we only have to prove that $\langle w, v\rangle = 0$. We have that

$$\begin{aligned}\langle w, v\rangle &= \left\langle u - \frac{\langle u,v\rangle}{\|v\|^2}v, v\right\rangle \\ &= \langle u, v\rangle - \left\langle \frac{\langle u,v\rangle}{\|v\|^2}v, v\right\rangle \\ &= \langle u, v\rangle - \frac{\langle u,v\rangle}{\|v\|^2}\langle v, v\rangle \\ &= \langle u, v\rangle - \frac{\langle u,v\rangle}{\|v\|^2}\|v\|^2 \\ &= 0\end{aligned}$$

Thus we are done. □

Proposition 7.1.5

Every orthonormal list of vectors are linearly independent.

Proof. Suppose that $v_1, \dots, v_n$ are orthnormal. We want to show that $v_n = \sum_{k=1}^{n-1} a_k v_k$ imnplies $a_1 = \cdots = a_{n-1} = 0$. Then

$$\langle v_n, v_i \rangle = \sum_{k=1}^{n-1} a_k (v_k \cdot v_i) = a_i \|v_i\|^2$$

for $i \in \{1, \dots, n-1\}$ since $v_1, \dots, v_{n-1}$ are orthonormal. But since $\langle v_n, v_k \rangle = 0$ we must have $a_i = 0$. This means that $a_1 = \cdots = a_{n-1} = 0$ and thus $v_1, \dots, v_n$ are linearly independent. □

7.2 Orthonormal Basis

Definition 7.2.1: Orthonormal Basis

A basis $v_1, \dots, v_n$ of an inner product space V with $\dim(V) = n$ is called orthonormal if

- $\|b_i\| = 1$ for $i \in \{1, \dots, n\}$
- $\langle b_i, b_j \rangle = \delta_{ij}$ for $i, j \in \{1, \dots, n\}$

Proposition 7.2.2

The orthonormal basis is indeed a basis for an inner product space V.

Proof. Since lists of orthonormal vectors are linearly independent and there are n vectors, they must also span V and thus is a basis. □

Theorem 7.2.3

Let $b_1, \dots, b_n$ be an orthonormal basis and $v = \sum_{k=1}^n a_k b_k$. Then

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

Proof. We have that

$$\begin{aligned}
\|v\|^2 &= \left\langle \sum_{k=1}^n a_k b_k, \sum_{k=1}^n a_k b_k \right\rangle \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j (b_i \cdot b_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \delta_{ij} \\
&= \sum_{k=1}^n |a_k|^2
\end{aligned}$$

and we are done. □

Proposition 7.2.4

Let $b_1, \dots, b_n$ be an orthonormal basis of V. Then

$$v = \sum_{k=1}^{n} \langle v, b_k \rangle b_k$$

Proof. Applying the inner product with b_i for each $i \in \{1, \dots, n\}$ gives $a_i = \langle v, b_i \rangle$ since $\langle b_i, b_k \rangle = 0$ for any $k \neq i$. Thus if $v = \sum_{k=1}^{n} a_k b_k$ then $v = \sum_{k=1}^{n} \langle v, b_k \rangle b_k$ and we are done. □

Theorem 7.2.5: Gram-Schmidt Procedure

Let $v_1, \dots v_m$ be a list of linearly independent vectors of V. Let $b_1 = \frac{v_1}{\|v_1\|}$. For $i = 2, \dots, m$. Define

$$b_i = \frac{v_i - \sum_{k=0}^{i-1} \langle v_i, b_k \rangle b_k}{\|v_i - \sum_{k=0}^{i-1} \langle v_i, b_k \rangle b_k\|}$$

Then $b_1, \dots, b_m$ are orthonormal and has the same span as $v_1, \dots, v_m$.

Theorem 7.2.6

Every finite dimensional inner product space has an orthonormal basis.

Proof. By the Gram-Schmidt procedure, every basis can be transformed into an orthonormal basis. □

7.3 Orthogonal Complements

Definition 7.3.1: Orthogonal Complement

Let U be a subset of an inner product space V. The orthogonal complement of U is defined as

$$U^\perp = \{v \in V | \langle v, u \rangle = 0 \text{ for all } u \in U\}$$

Proposition 7.3.2

Let U be a subset of V.

- U is a subspace of V if and only if $U^\perp$ is a subspace of V.
- $\{0\}^\perp = V$
- $V^\perp = \{0\}$
- $U \cap U^\perp = \{0\}$
- If $W \subseteq U$, then $U^\perp \subseteq W^\perp$

Proposition 7.3.3

Let U be a finite dimensional subspace of V. Then

$$U = (U^\perp)^\perp$$

Theorem 7.3.4

Suppose U is a finite dimensional subspace of V. Then

$$V = U \oplus U^{\perp}$$

and

$$\dim(U) + \dim(U^{\perp}) = \dim(V)$$

7.4 Orthogonal Maps

Definition 7.4.1: Orthgonal Maps

A linear map $T : V \to V$ is said to be orthogonal if

$$\langle T(v), T(w)\rangle = \langle v, w\rangle$$

for all $v, w \in V$.

One can think of orthgonal maps as orthgonality preserving maps. If $\langle v, w\rangle = 0$ then $\langle T(v), T(w)\rangle = 0$ which means orthogonality is preserved.

Proposition 7.4.2

Let $T : V \to V$ be a linear map over an inner product space V. Let A represent the linear map T. Then the following are equivalent.

- T is orthogonal
- A is orthogonal
- T maps orthonormal bases to orthonormal bases

Proof. Suppose that $T : V \to V$ is represented by A.

- (1) $\iff$ (2): We have that $\langle T(v), T(w)\rangle = v^T A^T Aw$. Thus it is equal to $v^T w$ if and only if $A^T A = I$.
- (1) $\iff$ (3): Suppose that T is orthogonal. Suppose that $\{b_1, \dots, b_n\}$ is orthonormal. Then $\langle T(b_i), T(b_j)\rangle = \langle b_i, b_j\rangle = 0$ for $i, j \in \{1, \dots, n\}$. Thus $\{T(b_1), \dots, T(b_n)\}$ is orthogonal. But they are also orthonormal since $\|T(b_i)\|^2 = \langle T(b_i), T(b_i)\rangle = \langle b_i, b_i\rangle = \|b_i\|^2 = 1$. This means that $\|T(b_i)\| = 1$ for $i \in \{1, \dots, n\}$.

  Now suppose that T maps orthonormal bases to orthonormal bases. Then if $\{b_1, \dots, b_n\}$ is orthonormal, we have

$$\begin{aligned}
\langle T(v), T(w)\rangle &= \left\langle \left(\sum_{k=1}^n v_k T(b_k)\right), \left(\sum_{k=1}^n w_k T(b_k)\right) \right\rangle \\
&= \sum_{k=1}^n (v_k w_k)\langle T(b_k), T(b_k)\rangle && (\langle T(b_i), T(b_j)\rangle = 0 \text{ if } i \neq j) \\
&= \sum_{k=1}^n v_k w_k \\
&= \langle v, w\rangle
\end{aligned}$$

  Thus we are done.

□

8 Orthgonality in $\mathbb{R}^n$

8.1 Reduction of Quadratic Forms over $\mathbb{R}$

Orthogonality is interesting for real matrices because the notion of similarity and congruence coinincide under orthogonality. Notice that being similar and congruent to a diagonal matrix at the same time means that there exists an invertible P such that $P^TAP = P^{-1}AP = D$.

In the remaining sections we treat the adjugate in the case of $\mathbb{R}^n$ and save the case for $\mathbb{C}^n$ in another chapter. Then V in the remaining sections will only denote $\mathbb{R}^n$.

Definition 8.1.1: Euclidean Vector Space

An Euclidean vector space is $\mathbb{R}^n$ equipped with an inner product.

Proposition 8.1.2

A function $b : V \times V \to \mathbb{R}$ is an inner product over $\mathbb{R}$ if and only if b is bilinear and positive definite.

Lemma 8.1.3: Polarization Identity

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

$$\langle x, y\rangle = \frac{1}{4}\|x+y\|^2 - \frac{1}{4}\|x-y\|^2$$

Proof. Simple proof using the fact that $\|x\|^2 = \langle x, x\rangle$. □

Proposition 8.1.4: Sylvester's Law of Inertia

A quadratic form q over $\mathbb{R}$ has the form

$$q(x_1, \dots, x_n) = \sum_{k=1}^{t} x_k^2 - \sum_{k=1}^{u} x_k^2$$

where $t + u = \operatorname{rank}(q)$. The pair (t, u) is called the signature of q. This reduced quadratic form is also unique in the sense that the number of positives and number of negatives of any two reduced forms are the same.

Moreover, every symmetric matrix is congruent to a matrix of the form

$$\begin{pmatrix} I_t & 0 & 0 \\ 0 & -I_u & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where (t, u) is the signature of the quadratic form.

Proof. We saw in theorem 1.2.7 that every quadratic form can be expressed as

$$q(y_1, \dots, y_n) = \sum_{k=1}^{n} c_k y_k^2$$

By doing a basis change with $b_k' = \frac{1}{\sqrt{c_k}} b_k$ whenever $c_k \neq 0$ will give us the above sum. For those that have $c_k = 0$, the terms vanish and are exactly in the kernel of the quadratic form thus $t + u = \operatorname{rank}(q)$.

The second part is direct from the fact that same quadratic forms with different matrix representations imply their representations are similar. □

8.2 Reduction of Inner Products

Definition 8.2.1: Dot Product

The dot product in $\mathbb{R}^n$ is defined to be the inner product given by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n$$

in standard basis.

Theorem 8.2.2

Let $\langle \cdot, \cdot \rangle$ be an inner product on a real vector space V. Then there exists an basis $b_1, \dots, b_n$ of V such that the inner product, when represented in the orthonormal basis, takes the form of exactly the dot product.

Proof. Let $\langle \cdot, \cdot \rangle$ be our inner product in question. Define a quadratic form by

$$q(x) = \langle x, x \rangle = \|x\|^2$$

We know that this quadratic form can be reduced to

$$q(x_1, \dots, x_n) = x_1^2 + \cdots + x_t^2 - x_{t+1}^2 - \dots x_{t+u}^2$$

Now we must have $u = 0$ since if $u > 0$, then the basis vector b_{t+1} satisfies $q(b_{t+1}) = -1$ and $q(b_{t+1}) = \langle b_{t+1}, b_{t+1} \rangle$ which is a contradiction since inner products are positive definite. Also $t = n$ since if $t < n$, then $\langle b_{t+1}, b_{t+1} \rangle = 0$ which is again a contradiction.

Using polarization, we see that $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$ in that basis and we are done. □

With this theorem, we know that any inner product can be expressed in the dot product as long as it is under a suitable basis. Thus we now reduce our discussion to only the dot product, as our standard inner product in $\mathbb{R}^n$.

The below theorem, while unrelated to the reduction of inner products, is a result of Gram-schmidt process that is only true for real matrices.

Theorem 8.2.3: QR Decomposition

Let A be an $n \times n$ matrix over $\mathbb{R}$. Then we can write $A = QR$ where Q is orthogonal and R is upper triangular.

Proof. We split the matrices into two cases. Firstly consider the case where A is invertible. We can treat A as a change of basis matrix from the basis $\{a_1, \dots, a_n\}$ where a_k is the column of A for $k \in \{1, \dots, n\}$. This change of basis matrix takes $\{a_1, \dots, a_n\}$ to $\{e_1, \dots, e_n\}$ which is the standard basis. Apply the Gram-schmidt process to $\{a_1, \dots, a_n\}$ to get $\{b_1, \dots, b_n\}$ which is an orthonormal basis. Let Q be the change of basis matrix from $\{b_1, \dots, b_n\}$ to $\{e_1, \dots, e_n\}$. Let R be the change of basis matrix from $\{a_1, \dots, a_n\}$ to $\{b_1, \dots, b_n\}$. Then clearly $A = QR$. We just have to show that Q is orthogonal and R is upper triangular.

Q being orthonormal is trivial since columns of Q are just $b_1, \dots, b_n$. Using the Gram-schimdt process, we can see that the change of basis from $\{b_1, \dots, b_n\}$ to $\{a_1, \dots, a_n\}$, each b_k is only affected by $a_1, \dots, a_k$ from the old basis. This means that the change of basis matrix must be upper triangular and its inverse must also be upper triangular.

Now we also have the case when A is not invertible. □

We now give an example of QR decomposition, in conjunction with the Gram-schmidt procedure.

Example 8.2.4

Find the QR decomposition of

$$A = \begin{pmatrix} -1 & 0 & -2 \\ 2 & 0 & -1 \\ 0 & -2 & -2 \end{pmatrix}$$

Proof. A quick check shows that A is invertible. Let a_1, a_2, a_3 be the columns of A. We start the Gram-schimdt process by taking the new basis $f_1 = \frac{a_1}{\|a_1\|} = \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix}$. We also need to keep track on the change of basis matrix. We have that $a_1 = \sqrt{5} f_1$.

For the next step, we find that

$$\begin{aligned} f_2 &= \frac{a_2 - (a_2 \cdot f_1) f_1}{\|a_2 - (a_2 \cdot f_1) f_1\|} \\ &= \frac{a_2}{\|a_2\|} \\ &= \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \end{aligned}$$

This means that $a_2 = 2f_2$.

Finally, we have that

$$\begin{aligned} f_3 &= \frac{a_3 - (a_3 \cdot f_1) f_1 - (a_3 \cdot f_2) f_2}{\|a_3 - (a_3 \cdot f_1) f_1 - (a_3 \cdot f_2) f_2\|} \\ &= \frac{a_3 - 2f_2}{\|a_3 - 2f_2\|} \\ &= \frac{a_3 - 2f_2}{\sqrt{5}} \\ &= \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \\ 0 \end{pmatrix} \end{aligned}$$

We have that $a_3 = 2f_2 + \sqrt{5} f_3$. Combining everything together, we have that

$$\begin{pmatrix} -1 & 0 & -2 \\ 2 & 0 & -1 \\ 0 & -2 & -2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{5}} & 0 & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & \sqrt{5} \end{pmatrix}$$

□

8.3 Adjoints

Proposition 8.3.1

Let V be an inner product space and $T : V \to V$ be a linear map. Then there exists a unique linear map $T^* : V \to V$ such that

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

for all $v, w \in V$.

Proof. Let T be a linear map. Then the function $\tau(v,w) = \langle T(v), w\rangle$ is a bilinear form since the inner product is bilinear. But we know that bilinear forms can be represented by a matrix multiplication, namely $\tau(v,w) = v^T Aw$ where A is defined as in theorem 1.3.2. Then treating Aw as the linear map $T^*(w)$ and since $v^T w = v \cdot w$, we have that $\tau(v,w) = v \cdot T^*(w)$ thus proving existence. Uniqueness follows naturally by construction of the matrix A. □

Definition 8.3.2: Adjoint of a Linear Map

T^* in the above case is called the adjoint of T.

Definition 8.3.3: Self-Adjoint

A linear map $T : V \to V$ is said to be self-adjoint if $T^* = T$

Proposition 8.3.4

Let T be a linear map represented by a matrix A. Then the following are true.

- T is self-adjoint if and only if A is symmetric.
- T is orthogonal if and only if $T^* = T^{-1}$.

Proof. Let T be self-adjoint. Then $Av \cdot w = v \cdot Aw$ for all $v, w \in V$. □

Proposition 8.3.5

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be self-adjoint. Then every eigenvalues of T are real.

Proof. Suppose that $T(v) = Av$ where A is a representation of T. Suppose that λ is an eigenvalue of T. Then $Av = \lambda v$ for some eigenvector $v \in \mathbb{R}^n$. Then taking complex conjugates give

$$\overline{Av} = \overline{\lambda v}$$
$$A\overline{v} = \overline{\lambda}\overline{v}$$

Taking the transpose of $Av = \lambda v$ gives $v^T A^T = \lambda v^T$ and $v^T A = \lambda v^T$. Multiplying $\overline{v}$ on bothe sides give

$$v^T A\overline{v} = \lambda v^T \overline{v}$$
$$\overline{\lambda} v^T \overline{v} = \lambda v^T \overline{v}$$

But $v^T\overline{v} = v_1\overline{v_1} + \cdots + v_n\overline{v_n} = |v_1|^2 + \cdots + |v_n|^2$ which is 0 if and only if v is 0. Since eigenvectors are taken to be nonzero, we must have $\lambda = \overline{\lambda}$ and thus λ is real. □

Theorem 8.3.6

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map on the inner product space $\mathbb{R}^n$ that is seldf-adjoint. Then there exists an orthonormal basis consisting of eigenvectors of T.

Equivalently, for every quadratic form q on V, there exists an orthonormal basis $b_1, \dots, b_n$ such that

$$q(x_1, \ldots, x_n) = \sum_{k=1}^{n} c_k x_k^2$$

for some $c_1, \ldots, c_n \in \mathbb{R}$.

Proof. Notice that the two statements are exactly the same and I will ommit the reason.

We prove by induction on n. Suppose that the theorem holds for $n-1$. Let T be the linear map. By the above we know that T has at least one eigenvalue in $\mathbb{R}$, say λ_1. Let f_1 be the corresponding eigenvector with magnitude 1.

Consider the orthogonal complement $W = \{w \in V | w \cdot f_1 = 0\}$ of f_1. Since W is the kernel of the linear map $S : V \to \mathbb{R}$ defined by $S(v) = v \cdot f_1$, it is a subspace of V dimension $n-1$. I claim that $T(W) \subseteq W$.

Let $w \in W$. We have

$$T(w) \cdot f_1 = w \cdot T(f_1) = w \cdot \lambda_1 f_1 = 0$$

by self-adjoint. Thus we have shown that $T(W) \subseteq W$.

Applying the induction hypothesis on W, we have an orthonormal basis $f_2, \ldots, f_n$ of W consisting of eigenvectors of T. By definition, $f_1, \ldots, f_n$ is an orthonormal basis and we are done. □

Notice that the above two statements are also equivalent to saying that every real symmetric matrix is congruent and similar to a diagonal matrix.

Proposition 8.3.7

If $T : \mathbb{R}^n \to \mathbb{R}^n$ is self-adjoint, and λ, μ are distinct eigenvalues of T with eigenvectors v, w, then $v \cdot w = 0$.

Proof. We have that

$$\begin{aligned} v^T A w &= v \cdot Aw \\ &= v^T \mu w \\ &= \mu(v \cdot w) \end{aligned}$$

and

$$\begin{aligned} v^T A w &= v^T A^T w \\ &= (Av)^T w \\ &= (\lambda v)^T w \\ &= \lambda v^T w \\ &= \lambda(v \cdot w) \end{aligned}$$

Comparing the two results, we have that $(\mu - \lambda)(v \cdot w) = 0$ and thus $v \cdot w = 0$. □

The proposition will prove itself to be useful in finding an orthonormal basis for self-adjoint linear maps.

8.4 Singular Value Decomposition

Theorem 8.4.1: Singular Value Decomposition for Linear Maps

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map of rank k. Then there eixsts unique positive numbers $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k \geq 0$ and orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ such that the matrix of T with respect to these bases is

$$\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where $D = \text{diag}(\gamma_1, \dots, \gamma_k)$. In fact, the γ_i are exactly the nonzero eigenvalues of T^*T, each one appearing as many times as the dimension of the corresponding eigenspace.

Theorem 8.4.2: Singular Value Decomposition for Matrices

Let $A_{m\times n}$ be a matrix. There exists unique singular values $\gamma_1 \geq \gamma_2 \geq \dots \gamma_k \geq 0$ where $k = \text{rank}(A)$, and orthogonal matrices P, Q such that

$$\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = P^T A Q$$

where $D = \text{diag}(\gamma_1, \dots, \gamma_k)$.

We present an example of singular value decomposition for illustration.

Example 8.4.3

Find the singular value decomposition of the matrix

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$$

Proof. Step 1: We compute the singular values of A, which is just the squareroot of the eigenvalues of A^TA. Now

$$A^TA = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}$$

We have that $c_{A^TA}(x) = x(360 - x)(90 - x)$. This means that the singular values are $\gamma_1 = \sqrt{360} = 6\sqrt{10}$ and $\gamma_2 = \sqrt{90} = 3\sqrt{10}$. Now we want P and Q such that

$$P^TAQ = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}$$

Step 2: We find the orthonormal eigenvectors of A^TA so that it forms the matrix Q. This gives

$$Q = \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Step 3: We now calculate P by finding the image of the above basis under A, and dividing it with the nonzero singular values. This gives

$$\begin{aligned} P &= \begin{pmatrix} \frac{1}{6\sqrt{10}}Ab_1 & \frac{1}{3\sqrt{10}}Ab_2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{pmatrix} \end{aligned}$$

This means that we are done. □