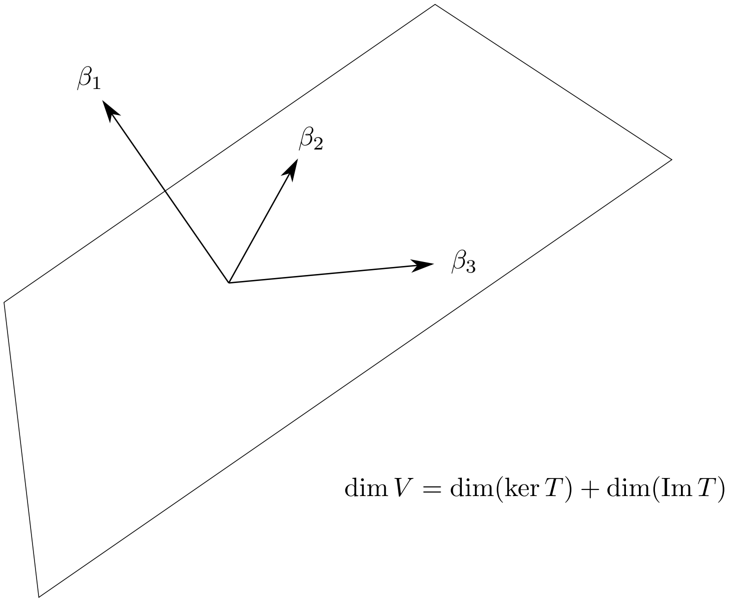
Notes on Linear Algebra

by Andreas Tsantilas
Professor Sylvia Serfaty

Spring 2020



$$\dim V = \dim(\ker T) + \dim(\operatorname{Im} T)$$

Contents

Introduction

This is a set of lecture notes, based on lectures given by Professor Sylvia Serfaty in the Spring of 2020. It is also based on the book *Linear Algebra Done Right* by Sheldon Axler, as well as *Linear Algebra* by Friedberg, Insel, and Spence.

Moreover, I have also interjected some of my own intuitions and explanations. For that reason, any errors the reader may encounter are most likely mine.

I hope that this is an engaging set of notes, and is a helpful supplement for a first course in linear algebra.

1 Vector Spaces

Linear algebra is the study of linear maps onto finite-dimensional vector spaces. The meanings of these two terms will be elucidated later on. For now, we will consider vector spaces over arbitrary fields (e.g., the real numbers or complex plane). Finally, we will define rigorously what we mean by subspace.

1.1 Review of Complex Numbers

A complex number is an ordered pair (a, b), where $a, b \in \mathbb{R}$. For convenience, we write complex numbers in the form

$$a + bi$$

Next, let's define the set of all complex numbers:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$$

Note that the dimension of $\mathbb{C}^n$ is $2n$ over the set $\mathbb{R}$. As you may already know, elements of $\mathbb{C}$ commute, associate, and are closed (that is, doing all operations give you an element in the same set). For $\alpha \in \mathbb{C}$:

1. $\forall \lambda \in \mathbb{C}, \lambda + 0 = \lambda, \lambda \cdot 1 = \lambda$ (additive and multiplicative identities).
2. $\forall \alpha, \exists! \beta \in \mathbb{C} : \alpha + \beta = 0$ (unique additive inverse).
3. If $\alpha \neq 0, \exists! \beta \in \mathbb{C} : \alpha\beta = 1$ (unique multiplicative inverse).
4. $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$.

Notationally speaking, we denote the additive inverse of α as $-\alpha$, and the multiplicative inverse as $1/\alpha$.

1.2 Arbitrary Fields

The stipulations above (commutativity, associativity, distributivity, existence of inverses, existence of identities) are part of a family of axioms called the ***field axioms***.

Notation ($\mathbb{F}$)**.** $\mathbb{F}$ denotes an arbitrary field (i.e., $\mathbb{F}$ follows the field axioms). $x \in \mathbb{F}$ is known as a "scalar."

A list of length n, $(x_1, x_2, \ldots, x_n)$, where $x_i \in \mathbb{F}$, is known as an n-tuple. In such tuples, order matters and repetitions have meaning, unlike in sets (Note: there are such things as 0-tuples!).

Definition 1.1 ($\mathbb{F}^n$)**.** $\mathbb{F}^n := \{(x_1, ...x_n) : x_i \in \mathbb{F} \text{ for } j = 1, \ldots, n\}$

That is $\mathbb{F}^n$ is the set of all n-tuples whose elements are in $\mathbb{F}$. Addition and subtraction are component-wise.

Definition 1.2 (0)**.** 0 is the n-tuple whose entries are all 0. Note that it is usually obvious in context what n is, so we can use the same notation in different dimensions.

One thing we should is to avoid using explicit coordinates whenever possible. Of course, sometimes it is very convenient to use them. But it is much cleaner to rely on group-theoretic notions that are more readily generalizable.

As in $\mathbb{F}$, we have addition and scalar multiplication in $\mathbb{F}^n$. There isn't really a useful way to construct multiplication that operates on two elements in $\mathbb{F}^n$ and outputs in $\mathbb{F}^n$.

For $x \in \mathbb{F}^n$, additive inverse is $-x \in \mathbb{F}^n$ such that $x + (-x) = 0$, where $0 \in \mathbb{F}^n$ is the identity element. Scalar multiplication is defined as

$$\lambda x = (\lambda x_1, ...\lambda x_n)$$

Where $\lambda \in \mathbb{F}$.

1.3 Definition of a Vector Space

The motivation behind defining a vector space comes from the properties of addition and scalar multiplication. More formally, a vector space is a set V that has addition and scalar multiplication, both of which commute, associate, have identities. Moreover, each $v \in V$ has an additive inverse $-v \in V$.

Definition 1.3 (Addition)**.** ***Addition*** on V is a function tht assigns an element $u + v \in V$ to each pair of $u, v \in V$.

Definition 1.4 (Scalar Multiplication)**.** ***Scalar Multiplication*** on V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbb{F}$ and each $v \in V$.

Definition 1.5 (Vector Space)**.** A ***vector space over*** $\mathbb{F}$ is a set V with addition and scalar multiplication, such that for $u, v, w \in V$ and for $\lambda_1, \lambda_2 \in \mathbb{F}$:

1. $u + v = v + u$; $\lambda_1\lambda_2 v = \lambda_2\lambda_1 v$ (commutative).
2. $(u + v) + w = u + (v + w)$; $\lambda_1(\lambda_2 v) = (\lambda_1\lambda_2)v$ (associative).
3. $0 \in V$ (additive identity).
4. $\forall v \in V, \exists! - v \in V : v + (-v) = 0$ (unique additive inverse).
5. $1 \in \mathbb{F}$ (existence of a multiplicative identity).
6. $\lambda_1(u + v) = \lambda_1 u + \lambda_1 v$ (distributive).

Definition 1.6 (Vector)**.** Elements of a vector space are called ***vectors***.

Notation. If S is a nonempty set, then $\mathbb{F}^S$ denotes the set of all functions from S to $\mathbb{F}$. The sum $f + g \in \mathbb{F}$ is defined by

$$(f + g)(x) = f(x) + g(x)$$

Similarly, scalar multiplication, $\lambda f \in \mathbb{F}^S$ is defined by

$$(\lambda f)(x) = \lambda f(x).$$

Is $\mathbb{F}^S$ a vector space? The answer is yes:

1. $0 \in \mathbb{F}^S$. 0 here is the zero function, such that $0(x) = 0$.

2. $\mathbb{F}^S$ is closed with respect to addition and scalar multiplication. That is, for $\lambda \in \mathbb{F}$ and $f, g \in \mathbb{F}^S$, $\lambda f + g \in \mathbb{F}^S$:

   $$(\lambda f + g)(x) = (\lambda f)(x) + g(x) = \lambda f(x) + g(x)$$

   Which is in $\mathbb{F}^S$.

1.4 Subspaces

Definition 1.7 (Subspace)**.** A subset U of V is called a ***subspace*** of V if U itself is a vector space.

The following is a test for a subspace:

1. $0 \in U$
2. For $u, w \in U$ and $a \in \mathbb{F}$, $\Rightarrow au + w \in U$.

Note that U is a subspace of V if and only if these properties are satisfied.

When dealing with vector spaces, it is useful to define an addition of spaces that results in another vector space.

Definition 1.8 (Sum of Subsets)**.** For $U_1, \ldots, U_m$ that are subsets of V, we define the ***sum*** to be

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}$$

Definition 1.9 (Direct Sum)**.** The sum $U_1 + \cdots + U_m$ is called a ***direct sum*** if each element of the sum can be written in only one way as a sum of $u_1 + \cdots + u_m$.

The sum of two subspaces U and W is a direct sum if and only if $U \cap W = 0$.

Notation (Direct Sum)**.** We denote the direct sum of subspaces $U_1, \ldots, U_m$ as

$$U_1 \oplus \cdots \oplus U_m$$

2 Finite-Dimensional Vector Spaces

A useful way to generate subspaces is to define a linear combination, which is the addition of scalar multiples.

Definition 2.1 (Linear Combination)**.** A ***linear combination*** of a set of vectors $\{v_1, \dots, v_n\}$ in V is a new vector of the form

$$a_1v_1 + \cdots + a_nv_n$$

Where $a_i \in \mathbb{F}$.

We can now define a neat way to generate a subspace of vectors, which follows from the definition of what a vector space is. That is to say, we leverage the fact that a vector space has to be "stable" under linear combination.

Definition 2.2 (Span)**.** The ***span*** is the set of all linear combinations of a set of vectors $\{v_1, \dots, v_n\}$ in V, which we denote via $\text{span}(v_1 \dots, v_n)$. Formally,

$$\text{span}(v_1, \dots, v_n) := \{a_1v_1 + \cdots + a_nv_n : a_1, \dots, a_n \in \mathbb{F}\}$$

When $\text{span}(v_1, \dots, v_n) = V$, we say the set $\{v_1, \dots, v_n\}$ ***spans*** (or ***generates***) V.

We are now equipped to tackle the idea of a finite-dimensional vector space, which is a core concept in linear algebra and the name of this section.

Definition 2.3. A vector space is ***finite-dimensional*** if some finite set of vectors spans (generates) the space.

One very widespread example of a vector space is the set of all polynomials whose coefficients are in $\mathbb{F}$.

Definition 2.4 (Polynomial)**.** A function $p : \mathbb{F} \to \mathbb{F}$ is a ***polynomial*** with coefficeints in $\mathbb{F}$ if there exist $a_0, \dots, a_m \in \mathbb{F}$ such that

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$$

We denote the set of all polynomials with degree less than or equal to m by

$$\mathcal{P}_m(\mathbb{F})$$

You should verify that $\mathcal{P}_m(\mathbb{F})$ is a finite-dimensional vector space. The set of polynomials that have no upper bound on degree is simply denoted $\mathcal{P}(\mathbb{F})$, and this is not a finite-dimensional vector space but an *infinite-dimensonial vector space.* Do you see how no finite list of vectors can generate the entire space?

2.1 Linear Independence

A crucial concept of finite-dimensional vector spaces is the notion of linear independence. This simple notion is the base of many proofs in the course.

Definition 2.5 (Linear Independence)**.** A set of vectors $\{v_1, \dots, v_m\}$ in V is said to be ***linearly independent*** if the only choice of $a_1, \dots, a_m \in \mathbb{F}$ that makes $a_1v_1 + \cdots + a_mv_m = 0$ is $a_1 = \cdots = a_m = 0$. In other words,

$$a_1v_1 + \cdots + a_mv_m = 0 \Rightarrow a_1 = \cdots = a_m = 0$$

We define the empty set $\varnothing$ to be linearly independent. This is useful for inductive proofs.

By the reasoning above, we deduce that a set is linearly independent if and only if each vector in the span has only one unique representation as a linear combination o v_i's. A set is linearly dependent if it's not linearly independent.

The lemma below is useful and appears in many proofs. It stipulates that given a linearly dependent set of vectors, we can throw out one vector without changing the span of the original set.

Lemma 2.1 (Linear Dependence)**.** Suppose $W = \{v_1, \dots, v_m\}$ is a linearly dependent list in V. Then there exists $j \in \{1, 2, \dots, m\}$ such that:

1. $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$
2. If we remove v_j from W, $\operatorname{span}(W \setminus \{v_j\}) = \operatorname{span}(W)$.

Let us continue with a statement about subsets of sets of vectors:

Lemma 2.2 (Subsets of Linearly Independent sets)**.** Let V be a vector space. For $S_1 \subseteq S_2 \subseteq V$, if S_1 is linearly dependent, then S_2 is linearly dependent.

Proof. Define S_2 to be the set of m vectors $S_2 = \{v_1, \dots, v_m\}$. Without loss of generality, let us say that S_1 is the subset of S_2 such that $S_1 = \{v_1, \dots, v_n\}$, where $n \leq m$. By definition of linear dependence, there exists $a_1, \dots, a_n$, not all zero, such that

$$a_1 v_1 + \cdots + a_n v_n = 0.$$

To prove that S_2 is linearly dependent as well, we observe that

$$a_1 v_1 + \cdots + a_n v_n + 0 v_{n+1} + \cdots + 0 v_m = 0$$

But not all coefficients $a_1, \dots, a_m$ are zero. Therefore S_2 is linearly dependent. ■

Corollary 2.3. Let V be a vector space. For $S_1 \subseteq S_2 \subseteq V$, if S_2 is linearly independent, then S_1 is linearly independent.

I would encourage you to figure out the proof for this on your own, as it is incredibly simple if you are familiar with propositional logic. A consequence of this lemma is the following "Replacement Theorem," from which lots of useful corollaries:

Theorem 2.1 (Replacement Theorem)**.** Let V be a vector space generated by a set G which contains n vectors, and L be a linearly independent subset of V with m vectors. Then

1. $m \leq n$, and
2. There exists a subset $H \subseteq G$ containing $n - m$ vectors such that $\operatorname{span}(L \cup H) = V$.

Proof. We begin with induction on m. Take $L = \varnothing$. Thus, $H = G$ yields the desired result.

Now suppose that the theorem holds for $m \geq 0$. Now we prove that the theorem is true for $m + 1$. Let $L = \{v_1, \dots, v_{m+1}\}$ be a linearly independent subset of V containing $m + 1$ vectors. We observe that the subset $\{v_1, \dots, v_m\}$ is also linearly independent by Lemma 2.2. By the induction hypothesis, we assume that $m \leq n$, and there is a subset

$\{g_1, \dots, g_{n-m}\}$ of G such that $\{v_1 \dots, v_m\} \cup \{g_1, \dots, g_{n-m}\}$ that generates V. Thus, there are scalars $a_1, \dots, a_m$ and $b_1, \dots, b_{n-m}$ such that

$$a_1 v_1 + \cdots + a_m v_m + b_1 g_1 + \dots b_{n-m} g_{n-m} = v_{m+1}$$

Note that $n - m > 0$, otherwise v_{m+1} would be a linear combinatoin of v_i's, which contradicts the fact that L is linearly independent. We know that some b_i, say, b_1, is nonzero. If not, then we could write v_{m+1} as a linear combination of v_i's, which contradicts the assumption that L is linearly independent.

This means we can express g_1 as a linear combination of $v_i's$ and g_i's:

$$g_1 = -\frac{1}{b_1}(a_1 v_1 + \cdots + a_m v_m + v_{m+1} + b_2 g_2 + \dots b_{n-m} g_{n-m})$$

Let $H = \{g_2, \dots, g_{n-m}\}$. Then $g_1 \in \text{span}(L \cup H)$. It follows that

$$\{v_1, \dots, v_m, g_1, \dots, g_{n-m}\} \subseteq \text{span}(L \cup H)$$

Because the set $\{v_1, \dots, v_m, g_1, \dots, g_{n-m}\}$ generates V (because it was our original generating set), and it is a subset of span$(L \cup H)$, then we know span$(L \cup H) = V$. Because H is a subset of G containing $n - m - 1 = n - (m + 1)$ vectors (because we removed g_1), the theorem is true for $m + 1$, completing the induction. ■

2.2 Bases and Dimension

We can now combine the concepts of spanning sets and linearly independent sets.

Definition 2.6 (Basis)**.** A ***basis*** of V is a set of vectors that is linearly independent and spans (generates) V.

Claim. A set of vectors $\{\beta_1 \dots, \beta_n\}$ is a basis of V if and only if every $v \in V$ can be written uniquely as a linear combination of the β_i's.

This follows from the fact that the β_i's are linearly independent (v is uniquely determined by the linear combination of basis vectors), and that the set is spanning (we can reach every $v \in V$).

From Theorem 2.3, we can extract our first important corollary:

Corollary 2.4 (Invariance of Cardinality)**.** Let V be a vector space with a finite basis. Then every basis for V contains the same number of vectors.

Proof. Suppose β is a basis that has n vectors, and let γ be another basis for V. If γ has more than n vectors, Because γ is linearly independent, and β generates V, we arrive at the conclusion that $m \leq n$, which is a contradiction. Therefore, we conclude the m vectors in γ satisfy $m \leq n$.

By symmetry, the same argument applies to β. We get the same constraint that $n \leq m$, so we conclude $m = n$. ■

This corollary implies something deep about vector spaces. The cardinality of any two basis sets are invariant. This invariant is how we define the dimension of a vector space.

Definition 2.7 (Dimension)**.** The ***dimension*** of a finite-dimensional vector space is the length of any basis of the vector space.

Notation. We denote the dimension of a vector space V by $\dim V$.

We can now freely use a "laziness principle" of sorts. If we know a basis of in question has the correct number of vectors in it (i.e., $\dim V$), then we only need to check that it is either linearly independent or spans V to verify that it is a basis.

Theorem 2.5 (Laziness Principle)**.** In order to prove a finite set of vectors $\{v_1, v_2, \ldots, v_n\}$ is a basis for V, it is sufficient to demonstrate any two of the three properties:

1. $n = \dim V$ (correct number of vectors).
2. $\{v_1, v_2, \ldots, v_n\}$ is linearly independent.
3. $\operatorname{span}(v_1, \ldots, v_n) = V$ (the vectors are generating).

That is, any two properties immediately imply the third.

Proof. Suppose $\dim V = n$, and $\{v_1, \ldots, v_n\}$ is linearly independent in V. By Theorem 2.3, we can complete this linearly independent set into a spanning one by taking the union with a subset H of G. However, H contains $n - n = 0$ vectors, so $\{v_1, \ldots, v_n\}$ is a basis.

Suppose now that $\{v_1, \ldots, v_n\}$ generates V. We know that there is some subset H that is basis of of V, because we can eliminate linearly dependent vectors of G without changing the span (Lemma 2.1). However, because all bases are the same length, and $|H| = |G|$, then $H = G$, so $\{v_1, \ldots, v_n\}$ is a basis. ■

We can now consider the inclusion-exclusion principle for summing subspaces:

Theorem 2.6 (Dimension of a Sum)**.** If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

Notice how if $U_1 + U_2$ is a direct sum, then $U_1 \cap U_2 = \{0\}$, and so the equation above takes the form

$$\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2.$$

Proof. Let $\{u_1, \ldots, u_m\}$ be a basis of $U_1 \cap U_2$, so $\dim(U_1 \cap U_2) = m$. Because it is a subset, u is linearly independent in U_1. By Theorem 2.3, we can complete it into a basis of U_1 by taking the union with some set H_1, of cardinality j. Therefore, we have $\dim U_1 = m + j$. Similarly, we take the union of u with a subset $\mathbb{H}_2$ of cardinality k to complete it into a basis containing $\dim U_2 = m + k$.

Next, let's show that $\dim(U_1 + U_2) = m + j + k$. We already know that $u \cup H_1 \cup H_2$ are generating sets. Now we must show that they are pairwise linearly independent to show that their union is a basis to show the dimension is $m + j + k$. It is obvious that u and H_1 are linearly independent, because of the mechanics of Theorem 2.3; similarly, u and H_2 are linearly independent.

[TO BE CONTINUED] ■

3 Linear Maps

We now turn to likely the most important topic of linear algebra.

3.1 The Vector Space of Linear Maps

Definition 3.1. A ***linear map*** from a vector space V to a vector space W is a function $T : V \to W$ with the following properties:

1. $T(u + v) = Tu + Tv$ for all $u, v, \in V$
2. $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{F}$ and all $v \in V$.

Another term for linear map is a ***linear transformation***.

Notation. The set of all linear maps from V to W is denoted

$$\mathcal{L}(V, W)$$

There are many examples of linear maps. They include the zero map, the identity map, differentiation, integration, and many, many more.

Next, we will prove the important fact that we can find a linear map that takes whatever values we wish on the vectors in a basis. The result below shows that a linear map is completely determined by its effects on a basis, which will be crucial in understanding the matrix representation of linear maps later on.

Theorem 3.1. Suppose $\{v_1, \ldots, v_n\}$ is a basis of V, and $\{w_1, \ldots, w_n\}$ is a basis of W. Then there exists a unique linear map $T : V \to W$ such that

$$Tv_j = w_j.$$

Proof. Define $T : V \to W$ by

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$$

Where $c_i \in \mathbb{F}$. For each j, set $c_j = 1$ and the other c's to be 0. This shows that it satisfies the property $Tv_j = w_j$ as desired. It is easy to show that T satisfies linearity, so this should be verified by the reader.

In order to show uniqueness, suppose $T \in \mathcal{L}(V, W)$ and that it satisfies the aforementioned property. Because T is linear, $T(c_jv_j) = c_jw_j$. By additivity, we see that

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n.$$

Therefore, T is uniquely determined on V by the equation above. ■

The results of this theorem imply that between two vector spaces with the same dimension, there is a unique linear map that can transform one basis into another, while preserving coordinates. This fact should not yet be apparent, as we have yet to define what coordinates are, and what coordinate representations of linear maps look like.

3.2 Algebra on $\mathcal{L}(V, W)$

Definition 3.2 (Addition and Scalar Multiplication on $\mathcal{L}(V, W)$)**.** For $S, T \in \mathcal{L}(U, V)$, and for $\lambda \in \mathbb{F}$, we define the sum $S + T$ to be

$$(S + T)(v) = Sv + Tv$$

And the product λT to be

$$(\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$.

As you should verify, $\mathcal{L}(V, W)$ is a vector space.

Previously, we have not defined a multiplication on a vector space (that is, we have not defined a function that takes in two elements of V and spits out another element in V). However, for linear maps, a product turns out to be quite useful. Therefore, we define the product of linear maps:

Definition 3.3 (Product of Linear Maps)**.** If $T \in \mathcal{L}(U, V)$ and $T \in \mathcal{L}(v, W)$, the product $ST \in T \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for $u \in U$.

As you should verify, this product of linear maps is associative ($(ST)R = S(TR)$), distributive ($(S + R)T = ST + RT$ and $S(R + T) = SR + ST$), and there exists an identity map I ($IT = TI = T$).

Warning. The product of linear maps is *not* commutative! That is, $ST = TS$ is not necessarily true.

One useful way to check if a map is indeed linear comes from the following theorem.

Theorem 3.2 (Linear Maps map 0 to 0)**.** Suppose $T \in \mathcal{L}(V, W)$. Then $T(0) = 0$.

Proof. $T(0) = T(0 + 0) = T(0) + T(0)$, by additivity. Add the additive inverse $-T(0)$ to both sides to get $T(0) = 0$. ■

3.3 Kernel and Image

In this section, we encounter two subspaces of V that are very closely connected to each linear map.

Definition 3.4 (Kernel)**.** For $T \in \mathcal{L}(V, W)$, the ***kernel*** of T, denoted $\ker T$, is the subset of V consisting of vectors that map to zero:

$$\ker T := \{v \in V : Tv = 0\}$$

Equivalently, some mathematicians will use the term “null space” instead of kernel.

Now an important question arises: Is the kernel of T a subspace of V? Indeed, it is.

Theorem 3.3 ($\ker T$ is a Subspace of V)**.** For $T \in \mathcal{L}(V, W)$, $\ker T$ is a subspace of V.

Proof. We already know that $0 \in \ker T$, by Theorem 3.2. Moreover, for $u, v \in \ker T$ and $\lambda \in \mathbb{F}$, we see that

$$T(\lambda u + v) = T(\lambda u) + T(v) = \lambda T(u) + T(v) = 0.$$

■

This next definition about functions in general is very closely related to the concept of the kernel.

Definition 3.5 (Injective)**.** A function $T : V \to W$ is called ***injective*** if $Tu = Tv$ implies $u = v$. Equivalently, the definition for injectivity is that $u \neq v$ implies $Tu \neq Tv$.

The gist of this definition is that T is injective if it maps distinct inputs to distinct outputs. No different inputs will yield the same output. For instance, The function $f(x) = x^2$ is *not* injective, because both 2 and -2 produce the same output of 4. For this reason, the phrase "one-to-one" is equivalently used by mathematicians. We shall see what exactly this property has to do with the kernel.

Theorem 3.4 (T is Injective $\equiv \ker T = \{0\}$)**.** Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\ker T = \{0\}$.

Proof. ($\Rightarrow$) Suppose T is injective. We know that T is a linear map, so $0 \in \ker T$ by theorem 3.2. Suppose $v \in \ker T$. Then $T(v) = 0 = T(0)$ By injectivity, we conclude that $v = 0$.

($\Leftarrow$) Suppose $\ker T = \{0\}$. Then we want to show that T is injective. Suppose $u, v \in V$, and $Tu = Tv$. we want to show $u = v$. We have $0 = Tu - Tv = T(u - v)$. Because $T(u - v)$ is in $\ker T$, and the only element is 0, then we deduce $u - v = 0$, so $u = v$, as desired. ■

Now, let's consider the outputs of a function.

Definition 3.6 (Image)**.** For $T : V \to W$, the ***image*** of T is the subset of W consisting of vectors that are of the form Tv for some $v \in V$:

$$\operatorname{Im} T = \{Tv : v \in V\}$$

Keep in mind that the kernel "lives" in the input space, and the image "lives" in the output space. Despite this, they are very deeply related. Like the kernel, the image is a subspace.

Theorem 3.5 ($\operatorname{Im} T$ is a Subspace of W)**.** If $T \in \mathcal{L}(V, W)$, then $\operatorname{Im} T$ is a subspace of W.

Proof. Because V is a vector space, $0 \in V$. And because $T(0) = 0$, we know that $0 \in \operatorname{Im} T$ (because 0 is $T(v)$ for some v, namely $v = 0$). Next, for $u, v \in \operatorname{Im} T$ and $\lambda \in \mathbb{F}$, $u = Tu'$, and $v = Tv'$. We have

$$T(\lambda u' + v') = \lambda Tu' + Tv' = \lambda u + v$$

Therefore, $u, v \in \operatorname{Im} T$ implies $\lambda u + v \in \operatorname{Im} T$. Therefore $\operatorname{Im} T$ is a subspace of W. ■

Much like the kernel, there is a property about functions in general that relates directly to the image of a linear map.

Definition 3.7 (Surjective)**.** A function $T : V \to W$ is called ***surjective*** if $\operatorname{Im} T = W$.

What this is essentially saying is that a function T is surjective if it "covers" every element of W. That is, every $w \in W = Tv$ for some $v \in V$. For this reason, mathematicians use the phrase "T is onto" to mean that T is surjective.

This next theorem is so important that it gets its own dramatic name.

Theorem 3.1 (Fundamental Theorem of Linear Maps)**.** Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\operatorname{Im} T$ is finite dimensional and:

$$\dim V = \dim(\ker T) + \dim(\operatorname{Im} T).$$

Remember this theorem well. It is the foundation for lots of application and theory alike.

Proof. Let $\{u_1, \dots, u_m\}$ be a basis for $\ker T$. This means that $\dim(\ker T) = m$. The set can be extended into a basis of V:

$$u_1, \dots, u_m, v_1, \dots, v_n$$

Therefore, $\dim V = m + n$. To complete the proof, let us show that $\dim(\operatorname{Im} T) = n$. With our basis of V, we have that any vector in V can be represented by a linear combination of u's and v's

$$w = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n$$

Where $a_i, b_i \in \mathbb{F}$. We apply T to both sides of the equation:

$$Tw = T(a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n)$$

However, because u_i is a basis for $\ker T$, those terms map to zero once linearity is applied. We are left with

$$Tw = b_1T(v_1) + \cdots + b_nT(v_n)$$

This last equation implies that $\{Tv_1, \dots, Tv_n\}$ spans $\operatorname{Im} T$. In order to show that Tv is a basis for the image, and therefore that $\dim(\operatorname{Im} T) = n$, we have to show linear independence. We know that v is a linearly independent set, that is not in $\operatorname{span}(u_1, \dots, u_m)$, because we used it to complete u into a basis for V. Thus, if

$$b_1T(v_1) + \cdots + b_nT(v_n) = 0$$

implies $b_1 = \cdots = b_n = 0$, then we will have confirmed linear independence. Indeed, this is the case. By linearity,

$$T(b_1v_1 + \cdots + b_nv_n) = 0$$

The argument of T must be zero. If not, then it would be another element in the kernel. Then some linear combination of u_i's would equal $b_1v_1+\cdots+b_nv_n$, but this cannot be since the two sets u and v are linearly independent. Therefore, the only option is that $b_1 = \cdots = b_n = 0$, so we have proved that $\{Tv_1, \dots, Tv_n\}$ is a basis for $\operatorname{Im} T$, so $\dim(\operatorname{Im} T) = n$, as desired. ■

As exercises, use the fundamental theorem of linear maps to prove the following:

Theorem 3.6. Suppose V and W are vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective (one-to-one).

Theorem 3.7. Suppose V and W are vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective (onto).

3.4 Matrices

By the conclusion we drew in Theorem 3.1, we know that if v is a basis of V, and $T \in \mathcal{L}(V, W)$, then the values of Tv determine the values of T on arbitrary vectors in V. Matrices are a way of neatly codifying this information. In order to rigorously define what a matrix is, we need the concept of an *ordered basis.*

Definition 3.8 (Ordered Basis)**.** Let V be a finite-dimensional vector space. An ordered basis for V is a basis endowed with a specific order.

Example. In $\mathbb{F}^3$, $\beta = \{e_1, e_2, e_3\} = \{(1,0,0), (0,1,0), (0,0,1)\}$ can be considered an ordered basis. Also, $\gamma = \{e_2, e_1, e_3\} = \{(0,1,0), (1,0,0), (0,0,1)\}$ is an ordered basis, but as ordered bases, $\beta \neq \gamma$.

For the vector space $\mathbb{F}^n$, we call $\{e_1, \dots, e_n\}$ the *standard ordered basis* or *canonical basis* for $\mathbb{F}^n$. For $\mathcal{P}_n(\mathbb{F})$, we call $\{1, x, x^2, \dots, x^n\}$ the standard ordered basis for $\mathcal{P}_n(\mathbb{F})$.

Definition 3.9 (Coordinates)**.** Let $\beta = \{\beta_1, \dots, \beta_n\}$ be an ordered basis for V. For $v \in V$, v is uniquely determined by some linear combination of the β_i's:

$$x = a_1\beta_1 + \cdots + a_n\beta_n$$

We define the coordinate vector of v relative to β:

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

Now that we have coordinates, we can go on to define what a matrix is.

Definition 3.10 (Matrix)**.** The *matrix* representation A of T in the ordered bases β, γ is an $m \times n$ array of of elements in $\mathbb{F}$. We write $A = [T]_\beta^\gamma$.

Note that the jth column of A is simply $[T(\beta_j)]_\gamma$. What this means is that we define a matrix based on what the linear transformation does to its basis vectors. It is not possible to have a matrix without specifying in which bases it is represented.

Definition 3.11 (Matrix Addition)**.** The sum of two matrices of the same size $m \times n$ is defined by adding the corresponding entries:

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} + \begin{bmatrix} b_{1,1} & \dots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \dots & b_{m,n} \end{bmatrix} = \begin{bmatrix} a_{1,1} + b_{1,1} & \dots & a_{1,n} + b_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} + b_{m,1} & \dots & a_{m,n} + b_{m,n} \end{bmatrix}$$

Next, we define scalar multiplication:

Definition 3.12 (Scalar Multiplication)**.** The product of a scalar and a matrix is simply the product of the scalar and each element:

$$\lambda \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \lambda a_{1,1} & \dots & \lambda a_{1,n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m,1} & \dots & \lambda a_{m,n} \end{bmatrix}$$

Representation of linear transformations follow the following properties, the proofs of which are left to the reader.

Remark. For a matrix representation of a linear transformation, the following properties hold:

1. $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$.
2. $[aT]_\beta^\gamma = a[T]_\beta^\gamma$.

As usual, matrices obey the same rules as linear transformations. After all, they are the same object in essence. As such, we must define a product, since there is such thing as a product for linear maps. The motivation for this multiplication makes it so that if $S : V \to W$ and $T : W \to Z$, where α, β, and γ are ordered bases for each space respectively:

$$[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta.$$

Therefore, we define the following to be matrix multiplication:

Definition 3.13 (Matrix Multiplication)**.** Suppoes A is an $m \times n$ matrix, and B is an $n \times p$ matrix. Then AB is the $m \times p$ matrix whose entry in row i, column j is given by:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

It is important to note that *matrix multiplication is not commutative.* That is, $AB \neq BA$ for most cases, even if both procucts are defined.

Theorem 3.8 (Product of Linear Maps in Matrix Form)**.** Let U, V, W be finite-dimensional vector spaces. Suppose $S : V \to W$ and $T : W \to Z$, where α, β, and γ are ordered bases for each space respectively:

$$[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta.$$

This is proven because this property was motivation for the definition of matrix multiplication.

Corollary 3.9. Let V be a finite-dimensional vector space with an ordered basis β, and $U, T \in \mathcal{L}(V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.

3.5 Invertibility and Isomorphisms

Let us begin this section by defining what we mean by invertible and an inverse.

Definition 3.14 (Matrix Inverse)**.** A linear map $T \in \mathcal{L}(V, W)$ is called *invertible* if there exists a map $T^{-1} \in \mathcal{L}(W, V)$ such that $T^{-1}T = I$.

Theorem 3.10 (Unique Inverse)**.** An invertible linear map has a unique inverse.

Theorem 3.2. (Invertible $\Leftrightarrow$ Injective and Surjective) A linear map is invertible if and only if it is injective and surjective.

Proof. Suppose $T \in \mathcal{L}(V, W)$, and that T is invertible, and $Tu = Tv$.

$$u = T^{-1}Tu = T^{-1}Tv = v$$

Because $u = v$, T is injective. In order to prove T is surjective, let $w \in W$:

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

Which shows that w is in the range of T for all w. Thus, $\operatorname{Im} T = W$, completing this direction.

Now, suppose T is injective and surjective. For each $w \in W$, let Sw be the unique element of V such that $TSw = w$. Clearly, TS is the identity map on W. Then

$$T(STv) = TS(Tv) = Tv$$

Therefore ST is the identitty map on V. To complete the proof, we simply prove S is linear:

$$T(aSw_1 + Sw_2) = TSaw_1 + TSw_2 = aTSw_1 + TSw_2 = aw_1 + w_2$$

Thus $Saw_1 + Sw_2$ is the unique element of V that T maps to $aw_1 + w_2$. This implies $S(aw_1 + w_2) = aSw_1 + Sw_2$, as desired. ■

Definition 3.15 (Isomorphism)**.** Two vector spaces are called *isomorphic* if there exists an invertible linear map between them.

Theorem 3.11 (Isomorphic ⇔ Same dimension)**.** Two finite dimensional vector spaces are isomorphic if and only if they have the same dimension.

Proof. First, assume the two spaces are isomorphic. By the fundamental theorem of linear maps:

$$\dim V = \dim \ker T + \dim \operatorname{Im} T$$

Because there is an invertible linear map, we know that $\dim \operatorname{Im} T = \dim W$. Moreover, because $\ker T = \{0\}$,

$$\dim V = \dim W$$

As desired.

Next, suppose V and W are finite dimensional and have the same dimension. Then define T such that on the bases v_i and w_i,

$$T(a_1v_1 + \cdots + a_nv_n) = a_1w_1 + \cdots + c_nw_n$$

T is well-defined, because v_i is a basis for V by Theorem 3.1. T is surjective because w_i spans W, and the kernel is $\{0\}$ because w_i are linearly independent. Therefore, V and W are isomorphic. ■

We can extend this theorem to familiar linear spaces:

Theorem 3.12 (Dimension of $\mathcal{L}(V, W)$)**.** Suppose V and W are finite-dimensional. Then $\dim \mathcal{L}(V, W) = \dim V \cdot \dim W$.

This follows from the fact that the dimension of a matrix is its rows times its columns, as it needs that many basis vectors to fully represent it in the canonical basis.

We will now turn to vectors that are linear maps from a vector space to itself.

Definition 3.16 (Linear Operator)**.** A linear map in $\mathcal{L}(V, V) = \mathcal{L}(V)$ is called an *operator*.

As you know, a linear map is invertible if and only if it is surjective and injective. However, for operators, all three conditions are equivalent!

Theorem 3.3. Let V be finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent:

1. T is invertible
2. T is injective
3. T is surjective

This can be proven by the all-powerful dimension theorem.

3.6 Change of Coordinates

Many areas of math use change of variables in order to greatly simplify a problem. Think of u-substitution in calculus. Similarly, in linear algebra, the question arises: how can we represent a coordinate vector in one basis to another? The idea is finding the identity represented in both bases.

Theorem 3.13. Let β, γ be two ordered bases for V, and let $Q = [I_V]_{\beta'}^{\beta}$. Then

1. Q is invertible.
2. For any $v \in V$, $[v]_\beta = Q[v]_{\beta'}$.

The proofs of these are rather trivial and will not be elaborated upon here.

Such a matrix Q is called the change of coordinate matrix; by (2) of the previous theorem, we can see that Q changes β' coordinates into β coordinates. Because Q is invertible, note that Q^{-1} changes β coordinates into β' coordinates.

Example. Let $\beta = \{(1, 1), (1, -1)\}$ and $\beta' = \{(2, 4), (3, 1)\}$. Since

$$(2, 4) = 3(1, 1) - 1(1, -1) \text{ and } (3, 1) = 2(1, 1) + 1(1, -1)$$

The corresponding change of basis matrix is

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

For instance,

$$[(2, 4)]_\beta = Q[(2, 4)]_{\beta'} = Q \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

We will, for the most part, only consider change of bases on linear operators:

Theorem 3.14. Let T be a finite-dimensional vector space. Let $T \in \mathcal{L}(V)$. Suppose $Q = [I_V]_{\beta'}^{\beta}$; that is Q changes β' coordinates to β coordinates. Then:

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

The proof of this follows from the fact that Q is the identity map, and that $T = IT = TI$, and from Theorem 3.9. The notion of changing bases gives rise to the idea of *similarity*:

Definition 3.17 (Similar Matrices)**.** Let A and B be matrices in $M_{n\times n}(\mathbb{F})$. We say that B is *similar* to A if there exists an invertible matrix Q such that

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

Observe that the relationship between similar matrices is symmetric, because the labels of Q and Q^{-1} are arbitrary.

3.7 Dual Spaces and Dual Maps

In linear algebra, there is a special place for linear maps from V to the scalar field $\mathbb{F}$; they have a special name.

Definition 3.18 (Linear Functional)**.** A *linear functional* on V is a linear map from V to $\mathbb{F}$. That is, it is an element of $\mathcal{L}(V, \mathbb{F})$.

That vector space also gets its own name:

Definition 3.19. The *dual space* of V, denoted V^*, is the vector space of all linear functionals on V. That is,

$$V^* = \mathcal{L}(V, \mathbb{F})$$

Theorem 3.15. Suppose V is finite-dimensional. Then V^* is also finite-dimensional and

$$\dim V^* = \dim V$$

This follows from the fact that the dimension of $\mathbb{F}$ is 1.
From theorem 3.1,

Definition 3.20. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis for V. Then the *dual basis* of β is the list $\varphi_1, \ldots, \varphi_n$ of elements of V^*, where

$$\varphi_i(v_j) = \delta_{ij}$$

Theorem 3.16 (Dual Basis is a Basis of the Dual Space)**.** Suppose V is finite dimensional. Then the dual basis of V is a basis of V^*.

Proof. Because there are n elements in both bases, we only need to prove linear independence. Suppose

$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0.$$

We know $(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_j) = a_j$. Therefore, because v_i are linearly independent, then $\varphi_1 \ldots a_n\varphi_n$ is lineraly independent.

■

Definition 3.21 (Dual Map, T^*)**.** If $T \in \mathcal{L}(V, W)$ then the *dual map* of T is the linear map $T^* \in \mathcal{L}(W^*, V^*)$ defined by

$$T^*(\varphi) = \varphi T \text{ for } \varphi \in W^*$$

We see that $T^*(\varphi)$ is defined to be the composition of linear maps φ and T (you should verify that such a composition is indeed possible). These dual maps obey certain algebraic properties:

Theorem 3.17 (Algebra of Dual Maps)**.** The following are algebraic properties of dual maps:

1. $(S+T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}(V, W)$.
2. $(\lambda T)^* = \lambda T^*$ for all $\lambda \in \mathbb{F}$.
3. $(ST)^* = T^* S^*$ for $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$.

Proof. The first two proofs are easy. For the third, suppose $\varphi \in W^*$. Then

$$(ST)^*(\varphi) = \varphi(ST)^* = (\varphi S)T = T^*(\varphi S) = T^*(S^*(\varphi)) = (T^* S^*)(\varphi)$$

Where the first, third, and fourth equalities hold by the definition of the dual map, and the second holds because composition of functions is associative, and the last from the definition of composition. ■

Next, our goal should be to describe $\ker T^*$ and $\operatorname{Im} T^*$ in terms of T. To do this, we need something called an annihilator.

Definition 3.22 (Annihilator U^0.)**.** For $U \in V$, the annihilator of U denoted U^0 is defined to be

$$U^0 = \{\varphi \in V^* : \varphi(u) = 0, \forall u \in U\}$$

The annihilator is a subset of the dual space ($U^0 \subseteq V^*$). In plain English, it is the set of all elements in the dual space such that φ of that element is 0. If φ sends all the elements in U to 0, then it is an element of the annihilator.

Theorem 3.18. U^0 is a subspace of V^*.

Clearly, $0 \in U^0$, where 0 is the zero linear functional on V. The proofs of the rest are left as exercises. We can now state some theroems about dual maps, by using the annihilator:

Theorem 3.19 (Dimension of U^0)**.** Suppose V is finite-dimensional and U is a subspace. Then

$$\dim U + \dim U^0 = \dim V.$$

This is proved by using the dimension theorem and the dual map of the inclusion operator, $i(u) = u$ for $u \in U$.

Theorem 3.20. Suppose V and W are finite-dimensional, and $T \in \mathcal{L}(V, W)$. Then

1. $\ker(T^*) = (\operatorname{Im} T)^0$
2. $\dim \ker(T^*) = \dim \ker(T) + \dim W - \dim V$

Theorem 3.21. Suppose V and W are finite-dimensional, and $T \in \mathcal{L}(V, W)$. Then

1. $\dim \operatorname{Im} T^* = \dim \operatorname{Im} T$
2. $\operatorname{Im} T^* = (\ker T)^0$

We will now define a useful idea in matrix manipulation, called the transpose.

Definition 3.23. The *transpose* of a matrix A, denoted A^t, is the matrix obtained by interchanginf the rows and columns of A. That is,

$$(a^t)_{ij} = a_{ji}$$

Theorem 3.22 (Transpose of Product)**.** $(AB)^t = B^t A^t$

This follows from the definition of transposition and matrix multiplication. Further still, note that the matrix of T^* is the transpose of T:

Theorem 3.23. Suppose $T \in \mathcal{L}(V, W)$. Then $[T^*]_\beta = [T^t]_\beta$.

4 Systems of Linear Equations

There are certain operations we can do on the rows and columns of matrices that preserve a certain property of the matrix's dimensionality.

4.1 Elementary Operations and Matrices

We define such operations below:

Definition 4.1 (Elementary Row/Column Operations)**.** Let A be an $m \times n$ matrix. Any one of the following three operations are considered elementary row or column operations:

1. Interchanging any two rows/columns of A
2. Multiplying any row/column by a nonzero scalar,
3. Adding any scalar multiple of a row/column of A to another row/column.

An elementary matrix is an $n \times n$ matrix obtained by performing either 1,2,or 3 on I_n. It is said to be "type" 1,2, or 3 based on the operation performed to obtain it. You should verify that if B is obtained from doing an elementary row/column operation on A, then there is an elementary matrix E which is $m \times m$ / $n \times n$ such that

$$B = EA$$

Theorem 4.1. Elementary matrices are invertible, and the inverse of an elementary matrix is an elementary matrix of the same type.

Below are some properties of these matrices:

1. Elementary matrices are always square
2. The inverse of an elementary matrix is an elementary matrix
3. E is an elementary matrix if and only if E^t is.

4.2 Rank and Inverse of Matrices

Definition 4.2 (Rank)**.** The *rank* of a linear map is defined to be the dimension of the range of the map. In other words,

$$\operatorname{rank} T := \dim(\operatorname{Im} T)$$

Equivalently, the rank of a matrix is the rank of the linear map it represents. Notice how $\operatorname{rank} T = \operatorname{rank}[T]_\beta^\gamma$

We need a result that allows us to perform rank-preserving operations on matrices:

Theorem 4.2 (Rank Preservation)**.** Let A be an $n \times n$ matrix. If P and Q are invertible $m \times m$ and $n \times n$ matrices respectively, then:

1. $\operatorname{rank} AQ = \operatorname{rank} A$
2. $\operatorname{rank} PA = \operatorname{rank} A$

3. $\operatorname{rank} PAQ = \operatorname{rank} A$

By this, it is clear that elementary row/column operations are rank-preserving, because they are simply matrices that are invertible. The following can be said about a matrix:

Theorem 4.3. The rank of any matrix equals the maximum number of linearly independent columns; the rank of a matrix is the subspace generated by its columns.

Now that we know how the rank is calculated, we can arrive at the following natural result:

Theorem 4.4. Let A be an $m \times n$ matrix of rank r. Then $r \leq m$ and $r \leq n$, and by means of a finite number of elementary row and column operations A can be transformed into the matrix

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

Where O is a zero matrix of a certain size. Thus, $D_{ii} = 1$ for $i \leq r$ and $D_{ij} = 0$ otherwise.

As a consequence, we can state the following corollary:

Corollary 4.5. Let A be an $m \times n$ matrix of rank r. Then there exist invertible matrices B and C, $m \times m$ and $n \times n$ respectively, such that

$$D = BAC$$

Where D is the matrix specified above.

This is true because in a finite number of elementary row and column operations, we can transform any matrix by Theorem 4.4; if we have a product of matrices $E_1 \cdots E_n$, the inverse would be $E_n^{-1} \cdots E_1^{-1}$ (this is obvious by multiplying the inverses on the left hand side). Yet another corollary may be stated:

Corollary 4.6. Let A be an $m \times n$ matrix. Then

1. $\operatorname{rank} A = \operatorname{rank} A^t$
2. $\operatorname{rank} A =$ maximum number of linearly independent rows (i.e., the dimension of the subspace created by the rows of the matrix).
3. The above is true for the columns of the matrix as well.

Proof. It is easy to see that if we do elementary row and column operations to obtain our matrix D, there is a symmetry between the transpose matrices; a row operation on one is a column operation on the other. From the corollary above, we know $D = BAC$; therefore $D^t = C^t A^t B^t$. Because B and C are invertible, so are the transposes. Therefore, the ranks are equal by Theorem 4.2. ■

Corollary 4.7. Every invertible matrix is a product of elementary matrices.

Proof. If A is an invertible matrix, then it is $n \times n$ and $D_A = I_n$. Therefore

$$I_n = BAC$$

Where B and C are the products $n \times n$ elementary matrices. Therefore,

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

Because C and B are the products of elementary matrices, their inverses is the product of inverse elementary matrices—which themselves are elementary matrices, as desired. ■

We are now able to state an important theorem:

Theorem 4.8. Let $T : V \to W$ and $U : W \to Z$ on finite-dimensional vector spaces. Let A, B be matrices such that their product is defined. Then:

1. $\operatorname{rank}(UT) \leq \min(\operatorname{rank} U, \operatorname{rank} T)$
2. $\operatorname{rank}(AB) \leq \min(\operatorname{rank} A, \operatorname{rank} B)$.

4.2.1 Inverse of a Matrix

Definition 4.3. Let A and B be $m \times n$ and $m \times p$ matrices, respectively. We define the *augmented matrix* $(A|B)$ to be the $m \times (n + p)$ matrices such that the first n columns are of A, and the remaining P columns are the columns of B.

When computing the inverse of a matrix, we augment it with the identity matrix $(A|I)$, and whatever row operations we do on A are reflected in the identiy. We do this until we have $(I|A^{-1})$, and the augmentation will be our desired inverse matrix. If this cannot be done, then A is not invertible.

Example. If we start with the matrix

$$(A|I) = \left[\begin{array}{ccc|ccc} 0 & 2 & 4 & 1 & 0 & 0 \\ 2 & 4 & 2 & 0 & 1 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 \end{array}\right]$$

And we attempt to use elementary row operations to transform the left hand side to I, we obtain

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/8 & -5/8 & 3/4 \\ 0 & 1 & 0 & -1/4 & 3/4 & -1/2 \\ 0 & 0 & 1 & 3/8 & -3/8 & 1/4 \end{array}\right]$$

Therefore,

$$A^{-1} = \begin{bmatrix} 1/8 & -5/8 & 3/4 \\ -1/4 & 3/4 & -1/2 \\ 3/8 & -3/8 & 1/4 \end{bmatrix}.$$

4.3 Systems of Linear Equations

We can use a matrix to neatly codify information we have been given from a linear system of equations. From a given set of equations, we can construct a so-called "coefficient matrix" with the coefficients of each system:

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix}$$

If we let

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and } b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Then we have the liberty to write the system simply as

$$Ax = b$$

A solution to the systemk is an n-tuple such that

$$As = b$$

The set of all solutions to the system is called the solution set. A system S is called *inconsistent* if the solution set is empty, and *consistent* otherwise.

Definition 4.4 (Homogeneous System of Linear Equations)**.** A system $Ax = b$ of linear equations in n unknowns is said to be *homogenous* if $b = 0$; otherwise, it is said to be *nonhomogeneous*.

Any homogenous linear system has the trivial solution of $x = 0$.

Theorem 4.9. Let $Ax = 0$ be a homogenous system of linear equations over $\mathbb{F}$. Let K denote the set of all solutions. Then $K = \ker L_A$, where L_A is the linear transformation represented by the matrix A.

This is apparent; if $As = 0$, then s is by definition in the kernel. If s is in the kernel, then it is a solution to $Ax = 0$. Therefore the solution set and the kernel are equal. This immediately produces the following important theorem about homogenous linear systems:

Theorem 4.10. Let $Ax = 0$ be a system of equations. Then the set of solutions forms a linear subspace. Moreover, if A is invertible, then the only solution is $x = 0$.

This is clear from the preceeding theorem, which states that the solution set is equal to the kernel of A, which is itself a subspace. If A is invertible, then $\ker A = \{0\}$, which is also a subspace. Therefore, for homogeneous linear systems, linear combinations of solutions are solutions as well.

We can state a similar theorem for nonhomogeneous systems:

Theorem 4.11. Let $Ax = b$ be a system of n linear equations in n unknowns. If A is invertible, then the system only has one solution given by

$$A^{-1}b.$$

Conversely, if the system only has one solution, then A is invertible.

5 Determinants

The determinant is a special scalar-valued function defined on the set of square matrices. It is n-linear, and posesses some interesting properties. Before delving into a more general version of determinants, we will first consider the simplest case for computing determinants, ignoring the set of 1×1 matrices.

5.1 Determinants of 2×2 Matrices

Definition 5.1 (Determinant of order 2)**.** Suppose we are given the matrix

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

Then we define the determinant of A, denoted by $\det A$ to be the scalar $ad - bc$.

Note that the determinant is not a linear map; $\det(A + B) \neq \det(A) + \det(B)$, as you should verify. The value of the determinant can actually provide very useful information on the nature of A and of L_A proper:

Theorem 5.1. Let $A \in \mathcal{M}_2(\mathbb{F})$. Then $\det(A) \neq 0$ if and only if A is not invertible. Moreover, if A is indeed invertible, then its inverse is given by

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

5.2 Determinants of Order n

In this section, we extend the definition of determinants to square matrices with $n > 2$. With this in mind, it is useful to cretate new notation in order to aid our generalizations:

Notation. Given $A \in \mathcal{M}_n(\mathbb{F})$, denote the $(n-1) \times (n-1)$ matrix made by deleting row i and column j by $\tilde{A}_{ij}$.

We are now able to compactly write the definition of the determinant for matrices.

Definition 5.2. Let $A \in \mathcal{M}_n(\mathbb{F})$. If $n = 1$, then we define $\det(A) = a_{1,1}$, or the only entry of A. For $n \geq 2$, we define $\det(A)$ recursively as:

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \cdot \det(\tilde{A}_{1j})$$

This definition is compliated, so the reader is encouraged to practice computing the determinant independently. When each row of a matrix is held fixed, we see that the determinant is a linear function of each row when the remaining rows are held fixed.

Theorem 5.2. The determinant of an $n \times n$ matrix is n-linear; that is, it is a linear function of each row when the remaining rows are held fixed. In other words,

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

where k is a scalar, and u, v, and a_i are row vectors in $\mathbb{F}^n$.

Corollary 5.3. If $A \in \mathcal{M}_n(\mathbb{F})$ has a row consisting entirely of zeros, then $\det(A) = 0$.

This follows from the fact that if a matrix is not invertible, then it can be reduced to a matrix with a row of zeros somewhere.

Theorem 5.4. The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is,

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

Theorem 5.5. If B is a matrix obtained by swapping any two rows of A, then $\det(B) = -\det(A)$.

Theorem 5.6. Let A be an $n \times n$ matrix and B be a matrix obtained by adding a scalar multiple of a row of A onto another row of A. Then $\det(B) = \det(A)$.

We can summarize the effects applying elementary matrices to A have:

1. If B is a matrix made by interchanging any two rows of A, then $\det(B) = -\det(A)$.
2. If B is a matrix made by multiplying a row of A by a nonzero scalar k, then $\det(B) = k\det(A)$.
3. If B is a matrix obtained by adding a scalar multiple of a row of A onto another row of A, then $\det(B) = \det(A)$.

5.3 Properties of Determinants

In this section, we summarize important properties of determinants.

Theorem 5.1 (Properties of determinants)**.** $\det(AB) = \det(A) \cdot \det(B)$.

$\det(A) \neq 0$ if and only if A is invertible.

If A is invertible, $\det(A^{-1}) = 1/\det(A)$.

$\det(A^t) = \det(A)$.

If A and B are similar matrices, then $\det(A) = \det(B)$.

5.4 Determinants, Abstractly

For the first part of this entire section, we have only considered the computational aspects of the determinant, and we have derived the above properties from said computations. It is now time to characterize the determinant by only three of these properties; that is, the only function $\delta : \mathcal{M}_n(\mathbb{F}) \to \mathbb{F}$ having these properties is the determinant. The first of these properties is that of being n-linear.

Definition 5.3 (n-linear function)**.** A function $\delta : \mathcal{M}_n(\mathbb{F}) \to \mathbb{F}$ is called an *n-linear* function if it is a linear function of each row of an $n \times n$ matrix when the remaining $n-1$ rows are held fixed; that is, for every $r = 1, \ldots, n$, we have

$$\delta \begin{pmatrix} a_1 \\ \vdots \\ u + kv \\ \vdots \\ a_n \end{pmatrix} = \delta \begin{pmatrix} a_1 \\ \vdots \\ u \\ \vdots \\ a_n \end{pmatrix} + k\delta \begin{pmatrix} a_1 \\ \vdots \\ v \\ \vdots \\ a_n \end{pmatrix}$$

where k is a scalar, and u, v, and a_i are vectors in $\mathbb{F}^n$.

This should look familiar, as we already mentioned that the determinant is an n-linear function. Next, we introduce the second property:

Definition 5.4 (Alternating function)**.** An n-linear function is called alternating if we have $\delta(x_1, \ldots, x_i, x_j, \ldots, x_n) = 0$ whenever $x_i = x_j$ (that is, two adjacent row vectors are identical).

With these two properties, we are allowed to state the following theorem about n-linear alternating functions.

Theorem 5.7. Let $\delta : \mathcal{M}_n(\mathbb{F}) \to \mathbb{F}$ be alternating and n-linear. Then

1. If B is a matrix obtained by interchanging two rows of A, then $\delta(B) = -\delta(A)$.
2. If A has two identical rows, then $\delta(A) = 0$.

Proof. Because δ is alternating and n-linear, we have

$$0 = \delta \begin{pmatrix} a_1 \\ \vdots \\ a_r + a_{r+1} \\ a_r + a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \delta \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ a_r + a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + \delta \begin{pmatrix} a_1 \\ \vdots \\ a_{r+1} \\ a_r + a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

$$= \delta \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \delta \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + \delta \begin{pmatrix} a_1 \\ \vdots \\ a_{r+1} \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \delta \begin{pmatrix} a_1 \\ \vdots \\ a_{r+1} \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

$$= 0 + \delta(A) + \delta(B) + 0$$

In the case that the two are not next to one another, we can perform $s - r$ interchanges in order to get the sequence

$$a_1, \ldots, a_s, a_r, a_{s+1}, \ldots, a_n$$

And then perform an additional $s - r - 1$ swaps to get the sequence

$$a_1, \dots, a_{r-1}, a_s, a_{r+1}, \dots, a_{s-1}, a_r, a_{s+1}, \dots, a_n$$

This means that we multiply $\delta(A)$ by -1 a total of $(s - r) + (s - r - 1)$ times. Because this is always odd, $\delta(B) = -\delta(A)$ as desired.

Next, for (2), we can make B the matrix obtained by swapping until the two identical rows are next to each other, so $\delta(B) = 0$. But we also know that $\delta(B) = -\delta(A)$, so $\delta(A) = 0$. ■

Corollary 5.8. If B is obtained by adding the multiple of one row of A onto another row, then $\delta(B) = \delta(A)$.

Proof. Like before, we can use linearity. If there is a linear combination of the rows $a_i + ka_j$, we observe that

$$\delta(B) = \delta(A) + k\delta(C)$$

Where C is the matrix with a_j in the ith row. But because C has two copies of the same row, $\delta(C)$ must be zero. ■

From this, we can clealy see the effect δ has on elementary matrices.

Corollary 5.9. Observe that

1. $\delta(E_1) = -\delta(I)$
2. $\delta(E_2) = k\delta(I)$
3. $\delta(E_3) = \delta(I)$.

Where E_1, E_2, E_3 are the three types of elementary matrices.

Theorem 5.10. Suppose $\delta(I) = 1$. Then we have $\delta(AB) = \delta(A)\delta(B)$.

Proof. If either A or B are not invertible, then the product AB is also not invertible. Therefore the determinant of the product and the product of the determinants both equal zero.

If they are both invertible, that means that A can be written as the product of elementary matrices. The effect an elementary matrix has on another matrix's δ is always multiplicative. Therefore, we can iterate over $A = E_1 \cdots E_m$, taking out each elementary matrix from the product, until we have $\delta(E_1) \cdots \delta(E_m)\delta(B)$, whereby we then collapse the product of deltas back into A, so we get $\delta(AB) = \delta(A)\delta(B)$. ■

Finally, we show the equivalence between the two functions.

Theorem 5.11. If $\delta : \mathcal{M}_n(\mathbb{F}) \to \mathbb{F}$ is an alternating n-linear function such that $\delta(I) = 1$, then $\delta(A) = \det(A)$ for every $A \in \mathcal{M}_n(\mathbb{F})$.

Proof. Clearly, if A is not invertible, $\delta(A) = \det(A) = 0$. If A is invertible, then it is the product of elementary matrices. Thus,

$$\delta(A) = \delta(E_1 \cdots E_m) = \delta(E_1) \cdots \delta(E_m) = \det(E_1) \cdots \det(E_m) = \det(E_1 \cdots E_m) = \det(A)$$

■

Having proved that δ yields the same results as the determinant for all invertible and non-invertible matrices, we can conclude that by just specifying these properties, $\delta(\cdot) = \det(\cdot)$.

6 Eigenvalues, Eigenvectors, and Diagonalization

6.1 Eigenvectors and Eigenvalues

Definition 6.1. Suppose $T \in \mathcal{L}(V)$. A scalar $\lambda \in \mathbb{F}$ is called an eigenvalue of T if there exists $v \in V$ such that $v \neq 0$ and

$$Tv = \lambda v$$

We can now state some equivalent conditions to being an eigenvalue.

Theorem 6.1 (Equivalent Conditions to Being an Eigenvalue)**.** The following conditions are equivalent for $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$:

1. λ is an eigenvalue of T;
2. $(T - \lambda I)$ is not injective (one-to-one);
3. $(T - \lambda I)$ is not surjective (onto);
4. $(T - \lambda I)$ is not invertible

We can derive these by manipulating the equation $Tv = \lambda v$ to get $(T - \lambda I)v = 0$. Now that we understand what an eigenvalue is, we can now define an eigenvector:

Definition 6.2. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$ is an eigenvalue of T. A vector $v \in V$ is called an eigenvector of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

We will now prove that one can have a basis of eigenvectors if there are n distinct eigenvalues.

Theorem 6.1. Let $T \in \mathcal{L}(V)$, and suppose $\lambda_1, \ldots, \lambda_n$ are distinct eigenvalues of T, and $v_1, \ldots, v_n$ are corresponding eigenvectors. Then $v_1, \ldots, v_n$ are linearly independent.

Proof. Suppose $v_1, \ldots, v_n$ are linearly dependent, and let k be the smallest number such that

$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

Therefore, there exist scalars a_i such that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}$$

If we apply T to both sides of this equation, we obtain

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}$$

Multiplying the above equation the last by λ_k and subtracting both, we get

$$0 = a_1(\lambda_k - \lambda_1)v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1})v_{k-1}$$

Because all the eigenvalues are distinct, we know that each of the a_i's must be equal to zero. But that means v_k is equal to zero. Becuase v_k is an eigenvector of T, this cannot happen. Therefore we must conclude that they are linearly independent as desired. ■

Corollary 6.2. Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Proof. This is follows from the fact that there can be a set of at most $\dim V$ linearly independent vectors, and distinct eigenvalues correspond to linearly independent eigenvectors. ■

6.2 Diagonal Matrices

Definition 6.3 (Diagonal Matrix)**.** An $n \times n$ matrix D is diagonal if it is of the form

$$\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

i.e., it only has entries on the main diagonal.

Definition 6.4. An operator $T \in \mathcal{L}(V)$ is called diagonalizable if there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix.

Equivalently, we observe that through change of basis, $[T]_\gamma$ is similar to a diagonal matrix:

$$[T]_\beta = Q^{-1}[T]_\gamma Q$$
$$[T]_\gamma = Q[T]_\beta Q^{-1}$$

Where $Q = [I]_\beta^\gamma$, or the matrix that changes β coordinates into γ coordinates.

Notice how if $\beta = \{v_1, \dots, v_n\}$ is an ordered basis for V such that $D = [T]_\beta$ is a diagonal matrix, we have $T(v_i) = \lambda v_i$, because $[v_i]_\beta$ will be the vector with a 1 in the ith slot, and so in multiplying we effectively isolate the scalar λ_i on the diagonal.

Conversely, if $\beta = \{v_1, \dots, v_n\}$ is an ordered basis for V such that $T(v_i) = \lambda v_i$, then clearly

$$[T]_\beta = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Theorem 6.2. If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable if and only if there exists an ordered basis β for T consisting of eigenvectors of T. Moreover, if T is diagonalizable, $\beta = \{v_1, \dots, v_n\}$ is an ordered basis of eigenvectors, and $D = [T]_\beta$, then D is a diagonal matrix and D_{ii} is the eigenvalue corresponding to v_i.

This theorem is the natural conclusion of our discussion thus far.

Theorem 6.3. Let $A \in \mathcal{M}_n(\mathbb{F})$. Then a scalar λ is an eigenvalue of A if and only if

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

Proof. This is because λ is an eigenvalue if and only if $(A - \lambda I)$ is not invertible, which is equivalent to saying $\det(A - \lambda I) = 0$. ■

Definition 6.5 (Characteristic Polynomial)**.** Let $A \in \mathcal{M}_n(\mathbb{F})$. The polynomial $f(t) = \det(A - tI)$ is called the characteristic polynomial of A.

From this, we observe that eigenvalues of A are simply the zeros of its characteristic polynomial. It is easily shown that similar matrices have the same characteristic polynomial, since they have the same determinants. Therefore, this definition can be extended to linear operators, irrespective of their basis. Try using the definition of characteristic polynomial, as well as what you know about polynomials, to prove that A can have at most n distinct eigenvalues.

Definition 6.6. A polynomial $f(t)$ in $\mathcal{P}(\mathbb{F})$ splits over $\mathbb{F}$ if there are scalars $c, a_1, \ldots, a_n$ such that it can be written as

$$f(t) = c(t - a_1) \cdots (t - a_n)$$

Theorem 6.3. The characteristic polynomial of any diagonalizable linear operator splits.

Proof. Let T be a diagonalizable linear operator on V, and let β be an ordered basis such that $[T]_\beta$ is a diagonal matrix. It is easy to see that

$$f(t) = \det(D - tI) = (\lambda_1 - t) \cdots (\lambda_n - t) = (-1)^n (t - \lambda_1) \cdots (t - \lambda_n)$$

■

From this, it is clear to see that if T does not have distinct eigenvalues, then there will be repeated zeros in the characteristic polynomial of T. Note that the converse is false; a polynomial that splits does not guarantee diagonalizability.

Definition 6.7. Let λ be an eigenvalue for an operator/matrix with characteristic polynomial $f(t)$. The multiplicity of λ is the largest positive integer k for which

$$(t - \lambda)^k$$

is a factor of $f(t)$.

If T is diagonalizable, then there is an ordered basis consisting of its eigenvectors. We know from Theorem (6.3) that $[T]_\beta$ has its eigenvalues along its diagonal. Therefore, because $f(t) = \det([T]_\beta - \lambda I)$, it is clear that an eigenvalue must manifest on the diagonal exactly as many times as its multiplicity. Thus, β contains as many linearly independent eigenvectors corresponding to the same eigenvalue as its multiplicity.

Definition 6.8. Let $T \in \mathcal{L}(V)$ with an eigenvalue of λ. The eigenspace of T corresponding to λ is defined as

$$E(\lambda, T) = \ker(T - \lambda I)$$

Theorem 6.4. Let $T \in \mathcal{L}(V)$, and let λ be an eigenvalue of T with multiplicity m. Then

$$1 \leq \dim E(\lambda, T) \leq m$$

Proof. We know that the eigenspace must be of demension at least one, otherwise there would be no vector $v \neq 0$ such that $(T - \lambda I)(v) = 0$, so by definition λ is not an eigenvalue of T.

For the other side of the inequality, pick a basis $\{v_1, \ldots, v_p\}$ for $E(\lambda, T)$, where p is the dimension of $E(\lambda, T)$. Complete it into a basis β for V. Let $A = [T]_\beta$. observe that for $1 \leq i \leq p$, v_i is an eigenvector of T corresponding to λ. Therefore, A can be written as:

$$\begin{bmatrix} \lambda I_p & B \\ O & C \end{bmatrix}$$

The characteristic polynomial is thus

$$f(t) = \det(A - tI) = \det((\lambda - t)I_p)\det(C - tI_{n-p}) = (\lambda - t)^p g(t)$$

Where $g(t)$ is some other polynomial. Therefore, the multiplicity of λ is at least p ($g(t)$ could theoretically have another factor of $(\lambda - t)$). However, $\dim(E(\lambda, T))$ is p. Thus

$$1 \leq \dim E(\lambda, T) \leq m$$

as desired. ■

These results lead to a crucial theorem in determining if a matrix is diagonalizable:

Theorem 6.4. Let $T \in \mathcal{L}(V)$ such that the characteristic polynomial of T splits. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then

1. T is diagonalizable if and only if the multiplicity of λ_i is equal to $\dim E(\lambda_i, T)$ for all i.
2. If T is diagonalizable, and β_i is a basis for $E(\lambda_i, T)$ for each i, then $\beta = \beta_1 \cup \cdots \cup \beta_k$ is an ordered basis for V consisting of eigenvectors of T.

Proof. Suppose T is diagonalizable. We know that $d_i = E(\lambda_i, T) \leq m_i$, where m_i is the multiplicity of λ_i. Let β be a basis of eigenvectors for V. For each i, let $\beta_i = \beta \cap E(\lambda, T)$, the set of vectors in β that are eigenvectors corresponding to λ_i. Clearly, the vectors in β_i are linearly independent in a space of dimension d_i. The cardinality of β_i, deonoted by n_i, is less than d_i for each i because β_i is a linearly independent subset of a subspace of dimension d_i, and $d_i \leq m_i$. The n_i's sum to n because β contains n vectors. Expressed as a sum,

$$n = \sum_{i=1}^{k} n_i \leq \sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} m_i = n$$

It follows that

$$\sum_{i=1}^{k} (m_i - d_i) = 0$$

But since $(m_i - d_i) \geq 0$ for all i, we conclude that $m_i = d_i$.

Conversely, assume the characteristic polynomial of T splits and that $m_i = d_i$. For each i, take an ordered basis β_i corresponding to $E(\lambda_i, T)$, the cardinalities of which are $d_i = m_i$. and let $beta = \beta_1 \cup \cdots \cup \beta_k$. We know that β is linearly independent, because eigenvectors corresponding to distinct eigenvalues are linearly independent. Furthermore, β contains

$$\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} m_i = n$$

vectors. Therefore, β is an ordered basis for V consisting of eigenvectors of V. ■

Theorem 6.5. Suppose $T \in \mathcal{L}(V)$, where V is finite-dimensional. suppose that $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_k, T)$$

is a direct sum. Furthermore,

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_k, T) \leq \dim V$$

Proof. Suppose

$$u_1 + \cdots + u_k = 0,$$

where each u_i is in $E(\lambda_i, T)$. Because eigenvectors corresponding to distinct eigenvalues are linearly independent, this implies that each $u_j = 0$ because each sum can only be written in one way as a sum of $u_1, \ldots, u_k$. ■

Lemma 6.1. If $V = W_1 \oplus \cdots \oplus W_k$, and β_i is a basis for W_i, then $\beta = \beta_1 \cup \cdots \cup \beta_k$ is a basis for V.

Now we can neatly summarize equivalent conditions to diagonalizability:

1. T is diagonalizable;
2. V has a basis consisting of eigenvectors of T;
3. The characteristic polynomial of T splits *and* for each eigenvalue of T, the multiplicity of λ equals $n - \operatorname{rank}(T - \lambda I)$
4. $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_k, T)$;
5. $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_k, T)$;

One important use for diagonalizing a matrix is computing matrix limits. For instance, if $A = PDP^{-1}$, then it becomes remarkably simple to compute any power of the matrix:

$$A^m = P^{-1} \begin{bmatrix} \lambda_1^m & & 0 \\ & \ddots & \\ 0 & & \lambda_k^m \end{bmatrix} P$$

Once we do that, we are able to more readily compute polynomials of matrices and of linear transformations:

$$p(A) = a_0 I + a_1 A + \cdots + a_n A^n$$

If A is diagonalizable, then

$$p(A) = P^{-1} \begin{bmatrix} p(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & p(\lambda_k) \end{bmatrix} P$$

In fact, we can define $f(A)$ for any f that is a power series that is convergent; e.g.,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

Which, if A is diagonalizable, is just

$$P^{-1}\begin{bmatrix} e_1^\lambda & & 0 \\ & \ddots & \\ 0 & & e_k^\lambda \end{bmatrix} P$$

Another use of diagonal matrices is for solving systems of differential equations.

6.3 Similarity Invariants

We have talked about how the similarity transformation

$$A \to P^{-1}AP$$

preserves many properties of A, in particular its eigenvalues, characteristic polynomial, and determinant. If two matrices are similar, it means they represent the same transformation in a different basis, which comes from the fact that

$$\det(P^{-1}AP) = \det(A)$$

from which we were able to compute the characteristic polynomial and eigenvalues.

Another important invariant is the so-called "trace" of a matrix. It is the function $Tr : \mathcal{M}_n(\mathbb{F}) \to \mathbb{F}$ defined by:

$$Tr(A) = \sum_{i=1}^{n} a_{ii}$$

Or the sum along the diagonals of a matrix. It is easy to show that $Tr(AB) = Tr(BA)$:

$$Tr(AB) = \sum_{i=0}^{n}\sum_{j=0}^{n} a_{ij}b_{ji}$$

$$Tr(BA) = \sum_{i=0}^{n}\sum_{j=0}^{n} b_{ij}a_{ji}$$

Without loss of generality, we are able to switch the i's and $j's$ to get

$$Tr(BA) = \sum_{j=0}^{n}\sum_{i=0}^{n} a_{ij}b_{ji}$$

As a consequence, we have

$$Tr(P^{-1}AP) = Tr(AP^{-1}P) = Tr(A)$$

So trace is another property of the matrix that is invariant under similarity transformations. It is independent from the choice of basis.

In particular, if the characteristic polynomial of A splits, we can see through judicious choice of P that:

$$Tr(A) = \sum_{i=0}^{n} \lambda_i$$

Where the $\lambda_i's$ are not necessarily distinct. Similarly, for the determinant, we have

$$\det(A) = \det(P^{-1})\det(D)\det(P) = \prod_{i=0}^{n} \lambda_i$$

This means that given any matrix at all, we are immediately able to find the sum and product of its eigenvalues by taking the trace and determinant, respectively, no change of basis needed. Because we have 2 equations, it becomes especially easy to solve for the eigenvalues for a 2×2 matrix, such that we simply need to solve the quadratic equation

$$x^2 - Sx + P = 0$$

Where S and P are the sum and product of the eigenvalues, respectively.

6.4 Matrix Limits and Markov Chains

In this section we are concerned with finding the limits of matrices:

$$\lim_{m \to \infty} A^m$$

Recall that if you have a scalar λ, its limits will be

Condition	Limit
$\lvert\lambda\rvert < 1$	0
$\lambda = 1$	1
$\lambda > 1$	$+\infty$
$\lambda \le -1$	no limit

As we saw before, the powers of a matrix are really dictated by the behavior of its eigenvalues in diagonal form; therefore, understanding the eigenvalues is key in understanding the limiting behavior of matrix powers.

When we iterate multiplications of A, it is called a dynamical system. The long-term behavior of such a system can be characterized as follows.

Theorem 6.6. Let $A \in \mathcal{M}_n(\mathbb{F})$ such that A is diagonalizable. The limit of

$$\lim_{m \to \infty} A^m$$

exists or converges if and only if for every eigenvalue, $-1 < \lambda_i \le 1$.

Proof. This follows from our computations involving powers of diagonal matrices. ■

Definition 6.9 (Probability Vector)**.** A column vector $\mathbf{x}$ is a probability vector if all its entries are greater than zero, and all of its entries sum to 1.

Definition 6.10 (Transition Matrix/Stochastic Matrix)**.** A matrix M is called a transition matrix if all its entries are nonnegative and each of its columns separately sum to 1.

Theorem 6.7. Let u be the vector whose entries are all 1, and M be a transition matrix. then

$$M^t u = u$$

In other words, U is an eigenvector of M whose eigenvalue is 1. This is also the case for M, because transpose matrices share the same characteristic polynomial and therefore eigenvalues.

Theorem 6.8. Let M be a transition matrix. Then:

1. 1 is always an eigenvalue of M
2. The rest of the eigenvalues satisfy $|\lambda| < 1$.

With these ideas in mind, we are able to study how linear systems evolve, and their long-term behavior. The transition matrix dictates how a given probability vector will evolve in the next step in the evolution; the probability vector gives the states of the system, with the corresponding probabilities in each slot of the vector (e.g., what proportion of a population lives in the city vs. the suburbs, if every year proportions of the population change their residences).

If we have an evolutionary process that only depends on the current state (and not on the time, earlier states, or other factors), then we have what is called a *Markov Process.* If, in particular, the number of states is finite, then we have what is known as a *Markov Chain.*

If for every step in the process, we evolve the probability vector by the transition matrix M, the long-term behavior of the system can be modeled by taking the limit of our initial state:

$$\lim_{k\to\infty} M^k \mathbf{x}_0 = \mathbf{x}_s$$

Where $\mathbf{x}_s$ is known as the steady-state distribution, such that

$$M\mathbf{x}_s = \mathbf{x}_s$$

if the limit exists, of course. Remarkably, it can be shown that the steady state does not at all depend on the initial conditions $\mathbf{x}_0$; only on the transition matrix!

More generally, dynamical systems involve nonlinear operations, but we still have to understand the eigenvalues of the linearized operators. For instance, if you have a map

$$\phi : \mathbb{R}^n \to \mathbb{R}^n$$

which is non linear, then the differential of the map at some steady-state $\mathbf{x}_s$ is linear:

$$\mathbf{D}\phi(\mathbf{x}_s).$$

By understanding the eigenvalues of that map, we can understand the stability of our state over time.

6.5 The Cayley-Hamilton Theorem

Definition 6.11 (Invariant Subspace)**.** Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called T-invariant if $u \in U$ implies $Tu \in U$.

You should verify that the following are all invariant subspaces:

1. $\{0\}$;
2. V;
3. $\ker T$;
4. $\operatorname{Im} T$;
5. $E(\lambda, T)$, for evert eigenvalue λ of T.

As we can see, a one-dimensional invariant subspace gives rise to the notion of an eigenvalue.

Definition 6.12. If a subset U of V is T-invariant, then we can define the *restriction operator* $T|_U \in \mathcal{L}(U)$ by:

$$T|_U(u) = Tu$$

for all $u \in U$.

Essentially, we are resetricting the domain of T to U; this remains a linear operator because no operations on u ever leave U. It is an operator in a lower-dimensional space.

In matrix form, if we have β be a basis for U completed into a basis for V, then we have a matrix of the form

$$[T]_\beta = \left[\begin{array}{c|c} T|_U & B \\ \hline O & C \end{array}\right]$$

Where the left part determines what T does to U, and the right part is the complement of U; the upper part is represented in the basis for U, and the lower part is represented in the basis we used to complete β. Notice how the lower-left part is necessarily 0; because U us T-invariant, it will always be able to be represented in the basis for U and cannot use basis vectors from the complement. In other words, it is upper triangular by blocks.

The ideal situation is to put the matrix in blocks such that

$$[T]_\beta = \left[\begin{array}{c|c} T|_U & O \\ \hline O & C \end{array}\right]$$

This can only be done if the complement of U is also T-invariant, which generally isn't the case.

Regardless, the effectiveness of this form is that the characteristic polynomial is simply

$$f(t) = \det(T - tI) = \det(T|_U - tI)\det(C - tI)$$

Which means

$$f(t) = f_{T|_U}(t)\det(C - tI)$$

Where $f_{T|_U}(t)$ is the characteristic polynomial of the operator T restricted to U. This lead us to a new theorem.

Theorem 6.9. If U is a T-invariant subspace of V, then the characteristic polynomial for $T|_U$, $f_{T|_U}(t)$, divides the characteristic polynomial of T.

Corollary 6.10. Any root of $f_{T|_U}(t)$ is also a root of $f(t)$, and every eigenvalue of $T|_U$ is an eigenvalue of T with multiplicity less than or equal to that for T.

Next, we move on to the notion of a cyclic subspace.

Definition 6.13. Let $T \in \mathcal{L}(V)$, and $x \in V$. Then the subspace

$$W = \operatorname{span}(x, T(x), T^2(x), \dots)$$

is called the T-cyclic subspace of V generated by x.

Lemma 6.2. Let W be the T-cyclic subspace of V generated by x. Then W is the smallest subspace of V that is T-invariant that contains x. That is, for U that is T invariant and contains x, $W \subseteq U$.

Proof. If U is T invariant and contains x, it must contain $T(x), T^2(x), \dots$; because the span is the smallest subset with all those vectors in it, then $W \subseteq U$. ■

Theorem 6.11. Let $T \in \mathcal{L}(V)$, and $x \in V$, and $k = \dim W \leq \dim V$ where W is the T-cyclic subspace of V generated by x. Then

1. $\{x, T(x), \dots, T^{k-1}(x)\}$ is a basis for W;
2. If
$$a_0 x + a_1 T(x) + \cdots + a_{k-1} T^{k-1}(x) + T^k(x) = 0,$$
then the characteristic polynomial of $T|_W$ is
$$f(t) = (-1)^k (a_0 + a_1 t + \cdots + a_{k-1} t^{k-1} + t^k)$$

Proof. 1. For $x \neq 0$, let j be the largest positive integer for which

$$\beta = \{x, T(x), \dots, T^{j-1}(x)\}$$

is linearly independent. Such a j must exist, because V is finite-dimensional. Let $Z = \operatorname{span}(\beta)$, so β is a basis for Z. Furthermore, $T^j(x) \in Z$, because it is linearly dependent on other vectors in β.

We show that Z is T-invariant. Let $w \in Z$:

$$w = b_0 x + \cdots + b_{j-1} T^{j-1}(x)$$

Applying T to both sides,

$$T(w) = b_0 T(x) + \cdots + b_{j-1} T^j(x)$$

Therefore, $T(w)$ is a linear combination of vectors in Z, so Z is T-invariant.

Furthermore, we know that because Z is T invariant and contains x, $W \subseteq Z$ from the lemma we proved above. However, it is obvious that $Z \subseteq W$, because W contains the first j powers of iterated T operations on x.

Therefore, $W = Z$. It follows that β is a basis for W, so

$$\dim(W) = j = k$$

as desired.

2. View β from above as an ordered basis for W. Let a_i be scalars such that

$$a_0x + \cdots + a_{k-1}T^{k-1}(x) + T^k(x) = 0$$

Observe that

$$[T|_W]_\beta = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{k-1} \end{bmatrix}$$

which has the characteristic polynomial

$$f(t) = (-1)^k(a_0 + a_1t + \cdots + a_{k-1}t^{k-1} + t^k)$$

as desired.

■

Next comes an incredibly important result:

Theorem 6.5 (Cayley-Hamilton)**.** Let $T \in \mathcal{L}(V)$, and let $f(t)$ be the characteristic polynomial of T. then

$$f(T) = 0$$

in other words, T is a solution to its own characteristic polynomial.

Proof. In order to do this, we must prove that $f(T)(x) = 0$ for all $x \in V$. Clearly, this occurs when $x = 0$.

For $x \neq 0$, let W be the T-cyclic subspace of dimension k generated by x. By the previous theorem, there exist scalars such that

$$a_0x + \cdots + a_{k-1}T^{k-1}(x) + T^k(x) = 0$$

Hence,

$$g(t) = (-1)^k(a_0 + \cdots + a_{k-1}t^{k-1} + t^k)$$

Is the characteristic polynomial of $T|_W$. Combining these two, we get

$$g(T)(x) = (-1)^k(a_0 + \cdots + a_{k-1}T^{k-1} + T^k)(x) = 0$$

We know that $g(t)$ divides $f(t)$ from Theorem 6.12; therefore T is a root of $f(t)$. ■

Theorem 6.12. Let $T \in \mathcal{L}(V)$, and suppose that $V = W_1 \oplus \cdots \oplus W_k$, where each W_i is a T invariant subspace. Then the characteristic polynomial of T is

$$\prod_{i=0}^{k} f_i(t)$$

where $f_i(t)$ is the characteristic polynomial of $T|_{W_i}$.

Proof. If we have these subspaces we can make a basis such that

$$\beta = \bigcup_{i=0}^{k} \beta_i$$

i.e., by uniting bases for each W_i. Therefore, the matrix is of the form

$$[T]_\beta = \left[\begin{array}{c|c|c} T|_{W_1} & O & O \\ \hline O & \ddots & O \\ \hline O & O & T|_{W_k} \end{array}\right]$$

Where each segment is in block form. The characteristic polynomial is simply the product of the determinants of the blocks minus tI—which is the characteristic polynomial of T. $\blacksquare$

7 Inner Product Spaces

Most applications of linear algebra involve some kind of measurement. In this section, we generalize the notion of length into a much richer theory of inner products.

7.1 Inner Products and Norms

We begin with the definition of the inner product:

Definition 7.1. An *inner product* on V is a function that takes each ordered pair (u, v) of elements in V to a scalar $\langle u, v\rangle \in \mathbb{F}$ and satisfies the properties for all $x, y, z \in V$ and all $\lambda \in \mathbb{F}$:

1. $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$.
2. $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$.
3. $\overline{\langle x, y\rangle} = \langle y, x\rangle$.
4. $\langle x, x\rangle > 0$ if $x \neq 0$.
5. $\langle x, x\rangle = 0$ if and only if $x = 0$.

Definition 7.2. An *inner product space* is a vector space V that is endowed with an inner product on V.s

Definition 7.3. (Canonical Inner Products) The canonical inner product for x, y in $\mathbb{R}^n$ and $\mathbb{C}^n$ is sometimes called the dot product, and is computed by

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y}_i$$

where x_i is the ith component, and we multiply it with the corresponding complex conjugate. Note in $\mathbb{R}^n$ this reduces to taking the product of each component.

You should verify that this is indeed an inner product defined on both $\mathbb{R}^n$ and $\mathbb{C}^n$.

Definition 7.4. Let $A \in \mathcal{M}_{m\times n}(\mathbb{F})$. We define the *adjoint* of A to be the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{A_{ji}}$ for all i, j (i.e., the conjugate transpose).

We can now define the canonical inner product for matrices:

Definition 7.5 (Frobenius Inner Product)**.** Let $V = \mathcal{M}_n(\mathbb{F})$, and define for $A, B \in V$

$$\langle A, B\rangle = Tr(B^* A)$$

This is an inner product defined on the space of $n \times n$ matrices.

From the definition of inner products, we can draw the following useful conclusions:

Theorem 7.1. Let V be an inner product space. Then, for all $x, y, z \in V$ and all $\lambda \in \mathbb{F}$, the following statements are true:

1. $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$

2. $\langle x, \lambda y\rangle = \overline{\lambda}\langle x, y\rangle$.

3. $\langle x, 0\rangle = \langle 0, x\rangle = 0$

4. If $\langle x, y\rangle = \langle x, z\rangle$, for all $x \in V$, then $y = z$.

Definition 7.6 (Orthogonality)**.** We say $x, y \in V$ are orthogonal with respect to the inner product on V if

$$\langle x, y\rangle = 0$$

The last statement of the above theorem is essentially equivalent to saying that if $\langle x, y\rangle = 0 \forall x \in V$, then $y = 0$, which is also equivalent to saying that the only vector that is orthogonal to all other vectors in V is 0.

Note that a vector space can be equipped with many inner products; there are an infinite number of them., so we must specify which one we are using. Sometimes, it's even beneficial to play with two inner products at the same time. However, different inner products give different notions of orthogonality.

Now, we will see how each inner product determines a norm:

Definition 7.7 (Norm)**.** Let V be an inner product space. For $x \in V$, the *norm* is defined by

$$\|x\| = \sqrt{\langle x, x\rangle}$$

Notice how for the canonical inner product on $\mathbb{R}^n$, the norm is simply the Euclidean distance from the origin of a certain vector.

Next, the nice properties we might be familiar with from Euclidean norms generalize:

Theorem 7.2. Let V be an inner product space over $\mathbb{F}$. Then, for all $x, y \in V$ and $\lambda \in \mathbb{F}$, the following hold:

1. $\|\lambda x\| = |\lambda| \cdot \|x\|$;

2. $\|x\| = 0$ if and only if $x = 0$. In any case, $\|x\| \geq 0$;

3. $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ (Pythagorean Theorem, for Real Spaces);

4. $|\langle x, y\rangle| \leq \|x\|\|y\|$ (Cauchy-Schwarz Inequality);

5. $\|x + y\| \leq \|x\| + \|y\|$ (Triangle Inequality).

Proof. Numbers (1) and (2) are obvious, and both follow from the definition of a norm.

The Pythagorean theorem is proved by definitions:

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

Cauchy-Schwarz is proved by the following:
If $y = 0$, then the equality is fulfilled. Assume $y \neq 0$; For any $\lambda \in \mathbb{F}$, we have

$$\begin{aligned}0 \leq \|x - \lambda y\|^2 &= \langle x - \lambda y, x - \lambda y\rangle \\ &= \langle x, x\rangle - \lambda\langle x, y\rangle - \overline{\lambda}\langle y, x\rangle + \lambda\overline{\lambda}\langle y, y\rangle\end{aligned}$$

If we set

$$\lambda = \frac{\langle x, y\rangle}{\langle y, y\rangle}$$

the inequality becomes

$$0 \leq \langle x, x\rangle - \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle}$$

from which (4) follows.

To prove the triangle inequality, we have

$$\begin{aligned}
\|x+y\|^2 &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \|x\|^2 + 2\Re(\langle x, y\rangle) + \|y\|^2 \\
&\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

where the fourth line follows from the Cauchy-Schwarz inequality. ■

For mathematicians, these properties (minus the Cauchy-Schwarz Inequality) constitute what defines a norm. In fact, a norm can even be defined without an inner product, but that is out of scope for now.

Definition 7.8 (Orthogonal Sets)**.** A subset S of V, where V is an inner product space, is said to be orthogonal if any two distinct vectors in S are orthogonal. A subset S of V is said to be orthonormal if S is orthogonal and $\|s\| = 1$ for all $s \in S$.

Equivalently, S is orthonormal if and only if $\langle s_i, s_j\rangle = \delta_{ij}$.

7.2 The Gram-Schmidt Procedure and Orthogonal Complements

Definition 7.9. Let V be an inner product space. Then a subset of V is an orthonormal basis for V if it is an ordered basis that is orthonormal.

Theorem 7.3 (Gram-Schmidt)**.** Let V be an inner product space, and $S = \{w_i, \ldots, w_n\}$ be a linearly independent subset of V. Let $S' = \{v_1 \ldots v_n\}$, where v_1 equal w_1 and

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} v_j$$

Then S' is an orthogonal set of nonzero vectors such that $\text{span}(S') = \text{span}(S)$. In particular, the set $\{\frac{v_1}{\|v_1\|}, \ldots, \frac{v_n}{\|v_n\|}\}$ is an orthonormal set.

Before we prove this, it is helpful to supply some intuitoin. What we do is we assign v_1 to w_1, and then project v_1 onto w_2; we then remove the orthogonal projection of w_2 that lies in the direction of v_1. In general, we remove the component of w_i that is orthogonally projected onto the span of the previous v's.

Lemma 7.1. The coordinates of a vector $y \in V$ in an ordered orthogonal basis β of V is

$$y_i = \frac{\langle y, \beta_i\rangle}{\|\beta_j\|^2}$$

where y_i is the ith entry, and β_i is the ith orthonormal basis vector.

Proof. Because $y \in \operatorname{span}(\beta)$, y can be written as a linear combination of β_i's:

$$y = \sum a_i \beta_i$$

To isolate each a_j, we simply take the product with each v_j:

$$\langle y, \beta_j \rangle = \langle a_1 \beta_1, \beta_j \rangle + \cdots + \langle a_n \beta_n, \beta_j \rangle$$

However, for $i \neq j$, the β_i's are orthogonal to β_j; therefore, we are only left with

$$a_j \langle \beta_j, \beta_j \rangle = a_j \|\beta_j\|^2$$

■

Corollary 7.4. An orthogonal set is linearly independent.

It becomes apparent that in the Gram-Schmidt procedure, we can see that we are essentially removing the coordinates that have already been expressed. We may now prove the Gram-Schmidt procedure.

Proof. The proof proceeds by induction on n. If $n = 1$, then clearly the S is linearly independent, because $w_1 \neq 0$. Setting $v_1 = w_1$, the theorem holds.

Assume it holds up to $n - 1$. $S'_{k-1} = \{v_1, \ldots, v_{k-1}\}$ is an orthonormal set. If $v_k = 0$, then that implies that $w_k \in \operatorname{span}(S'_{k-1}) = \operatorname{span}(S_{k-1})$, which can't occur because S_k was taken to be a linearly independent set.

It follows that

$$\begin{aligned}\langle v_k, v_i \rangle &= \langle w_k, v_i \rangle - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} \langle v_j, v_i \rangle \\ &= \langle w_k, v_i \rangle - \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} \langle v_i, v_i \rangle \\ &= 0\end{aligned}$$

Because v_k is orthogonal to all v_i, where $i < k$, then S'_k is an orthogonal set.

Now we show that $\operatorname{span}(S'_k) = \operatorname{span}(S_k)$. As we have seen before, the v_k's are in the span of the w_k's, so

$$\operatorname{span}(S'_k) \subseteq \operatorname{span}(S_k)$$

However, from corollary 7.4, we know that S'_k is a linearly independent set with k vectors. Therefore

$$\dim(\operatorname{span}(S'_k)) = \dim \operatorname{span}(S_k) = k$$

Therefore

$$\operatorname{span}(S'_k) = \operatorname{span}(S_k)$$

as desired. ■

Corollary 7.5. Every finite-dimensional inner product space has an orthonormal basis.

Corollary 7.6. Let V be a finite-dimensional inner product space, with an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$. Let T be a linear operator on V, and let $A = [T]_\beta$. Then

$$A_{ij} = \langle T(v_j), v_i \rangle.$$

Definition 7.10. Let β be an orthonormal subset (possibly infinite) of an inner product space V, and let $x \in V$. We define the Fourier Coefficients of x relative to β to be the scalars $\langle x, y\rangle$ where $y \in \beta$.

This is most commonly used for functions on an interval with the inner product

$$\int_0^{2\pi} f(x)\overline{g}(x)dx$$

onto the basis

$$(\cos nt, \sin nt)$$

or, more generally,

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}dx$$

for a function f.

Definition 7.11. Let S be a nonempty subset of an inner product space V. We dedfine $S^\perp$ (read "S perp") to be the set of all vectors in V that are orthogonal to every vector in S.

$$S^\perp = \{x \in V : \langle x, y\rangle = 0, \forall y \in S\}$$

You should verify that $S^\perp$ is a linear subspace.

Theorem 7.7. Let W be a finite-dimensional subspace of the inner product space V. Then $\forall y \in V$, there exist unique vectors $u \in W$ and $z \in W^\perp$ such that $y = u + z$.

This means that $V = W \oplus W^\perp$; moreover, if $\{v_1, \ldots, v_k\}$ are an orthonormal basis for W, then

$$u = \sum_{j=0}^{k} \langle y, v_i\rangle v_i$$

Proof. Let $\{v_1, \ldots, v_k\}$ be an orthonormal basis for W, and let u be defined in the preceeding manner, and let $z = y - u$. Clearly $u \in W$ and $y = u + z$.

To show that $z \in W^\perp$, it suffices to show that z is orthogonal to each v_j.

Via computation, we see that

$$\langle z, v_j\rangle = 0$$

To show uniqueness, we prove in the standard manner. This completes the proof. ■

Corollary 7.8. u is the unique vector in W that is "closest" to y; that is, for any $x \in W$,

$$\|y - x\| \geq \|y - u\|$$

And this is an equality if and only if $x = u$.

Theorem 7.9. Suppose that $S = \{v_1, \ldots, v_k\}$ is an orthonormal set for an n-dimensional inner product space V. Then:

1. S can be extended into an orthonormal basis $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for V;
2. If $W = \text{span}(S)$, then $S_1 = \{v_{k+1}, \ldots, v_n\}$ is an orthonormal basis for $W^\perp$.
3. If W is any subspace of V, then $\dim V = \dim W + \dim W^\perp$.

Proof. (1) can be proven by the completion theorem, and then it can be made into an orthonormal set via Gram-Schmidt.

(2) Trivial.

(3) Recall that $V = W \oplus W^\perp$, so the dimensions are additive. ■

7.3 Adjoint of a Linear Operator

Theorem 7.10. Let V be a finite-dimensioal inner product space over $\mathbb{F}$, and let $g : V \to \mathbb{F}$ be a linear transformation. Then there exists a unique vector $y \in V$ such that

$$g(x) = \langle x, y \rangle$$

for all $x \in V$.

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V, and let

$$y = \sum \overline{g(v_i)} v_i$$

Define $h(x) = \langle x, y \rangle$, which is clearly linear. Furthermore, we have

$$h(v_j) = g(v_j)$$

Since h and g both agree on β, and they are both linear, and any vector can be expressed as a linear combination of basis vectors, we have $g = h$.

Uniqueness is proven in the traditional manner. ■

Theorem 7.11. Let V be a finite-dimensional inner product space and let $T \in \mathcal{L}(V)$. Then there exists a unique function $T^* : V \to V$ such that

$$\langle T(x), y \rangle = \langle x, T^*(y) \rangle$$

for all $x, y \in V$.

Proof. Let $y \in V$, define the functional $g : V \to \mathbb{F}$ by

$$g(x) = \langle T(x), y \rangle$$

for all $x \in V$. First we show that g is linear, done by the usual method.

We know from theorem 7.10 that we can obtain a unique vector $y' \in V$ such that $g(x) = \langle x, y' \rangle$; that is, $\langle T(x), y \rangle = \langle x, y' \rangle$ for all $x \in V$. Defining $T^* : V \to V$ by $T^*(y) = y'$, we have

$$\langle T(x), y \rangle = \langle x, T^*(y) \rangle$$

as desired. Linearity can be proven in the usual manner. ■

Theorem 7.12. Let V be a finite-dimensional inner product space and let β be an orthonormal basis for V. If T is a linear operator on V, then

$$[T^*]\beta = [T]^*_\beta$$

Where $*$ denotes conjugate transposition for a matrix.

Theorem 7.13. The following properties hold for both linear operators and matrices:

1. $(T + U)^* = T^* + U^*$;
2. $(cT)^* = \bar{c}T^*$ for any $c \in \mathbb{F}$;
3. $(TU)^* = U^*T^*$;
4. $T^{**} = T$;
5. $I^* = I$.

7.4 Normal and Self-Adjoint Operators

Lemma 7.2. Let $T \in \mathcal{L}(V)$, and V be a finite-dimensional inner product space. If T has an eigenvector, so does T^*.

Proof. Suppose that v is an eigenvector with corresponding eigenvalue λ. Then, for any $x \in V$;

$$0 = \langle 0, x\rangle = \langle (T - \lambda I)(v), x\rangle = \langle v, (T - \lambda I)^*(x)\rangle = \langle v, (T^* - \overline{\lambda} I)(x)\rangle = 0$$

Hence, v is orthogonal to the range of $T^* - \overline{\lambda} I$, so it is not onto nd hence not one-to-one. Therefore there exists an eigenvector with corresponding eigenvalue $\overline{\lambda}$. ■

Theorem 7.1 (Schur's Theorem)**.** Let $T \in \mathcal{L}(V)$, where V is a finite-dimensional inner product space. Suppose the characteristic polynomial of T splits; then there exists an orthonormal basis β for V such that $[T]_\beta$ is upper triangular.

Proof. The proof proceeds by mathematical induction on $n = \dim V$. The result is immediate if $n = 1$.

Suppose, then, that Schur's theorem holds for $(n-1)$-dimensional inner product spaces whose characteristic polynomials split. By the lemma, we assume that T^* has a unit eigenvector z. Suppose that $T^*(z) = \overline{\lambda} z$ and that $W = \operatorname{span}(z)$. We show that $W^\perp$ is T-invariant. If $y \in W^\perp$, and $x = cz \in W$, then

$$0 = \langle y, z\rangle = \langle y, T^*(z)\rangle = \langle T(y), z\rangle = 0$$

Therefore, $T(y)$ is in $W^\perp$, so it is T-invariant.

We know that $\dim W = 1$, so $\dim W^\perp = n - 1$. From what we have seen before, the characteristic polynomial of $T|_{W^\perp}$ divides the characteristic polynomial of T; by what we assumed by induction, the characteristic polynomial of an $(n-1)$-dimensional space splits. We find a basis γ such that $T|_{W^\perp}$ is upper triangular.

Let $\beta = \gamma \cup \{z\}$. In this orthonormal basis, $[T]_\beta$ is upper triangular, as desired. ■

Definition 7.12. We say an operator T is normal if

$$TT^* = T^*T$$

and similarly for matrices.

In general, this is not the case.

Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation by θ, where $0 < \theta < \pi$. The matrix representation in the canonical basis is:

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Note that $AA^* = I = A^*A$; it normal. Moreover, the adjoint of a rotation operator is its own inverse.

Example. An interesting example is that of the symmetric matrix, such that $A = A^t$; over the real numbers, such a matrix equals its adjoint; this is known as a hermitian operator and it is normal.

Another interesting case is for real skew-symmetric matrices; that is, $A^t = -A$. Then A is normal because both AA^t and A^tA are equal to $-A^2$.

Theorem 7.2. The following are properties of normal operators:

1. $\|T(x)\| = \|T^*(x)\|$ for all $x \in V$;
2. $T - \lambda I$ is normal for all $c \in \mathbb{F}$;
3. If x is an eigenvector of T with eigenvalue λ, then x is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$.
4. If λ_1 and λ_2 are distinct eigenvalues of T with corresponding eigenvectors x_1 and x_2, then x_1 and x_2 are orthogonal.

Theorem 7.3. Let $T \in \mathcal{L}(V)$, and V be a finite-dimensional complex inner product space. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.

That is, T can be diagonalized in an orthonormal basis of eigenvectors.

Proof. If T can be diagonalized in an orthonormal basis β, then $[T^*]_\beta$ has the conjugate eigenvalues on the diagonals. Therefore, the product of the matrices commute; moreover,

$$[T^*T]_\beta = [TT^*]_\beta$$

Therefore T is normal.

Convresely, if T is normal in a complex inner product space, then the characteristic polynomial of T must split. By Schur's theorem, there is an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$ such that $[T]_\beta$ is upper triangular. In particular, the first column of $[T]_\beta$ is an eigenvalue; $[T]_\beta(v_1) = [T_{11}]_\beta$. Let $k \leq n$ be the maximal number such that $\{v_1, \ldots, v_{k-1}\}$ are eigenvectors of T.

For all $j \leq k-1$, $T(v_j) = \lambda_j v_j$, so $T^*(v_j) = \overline{\lambda_j} v_j$, so v_j is also an eigenvector for T^*. Let us compute $T(v_k)$:

$$T(v_k) = \sum_{j=1}^{k} \langle T(v_k), v_j \rangle v_j$$

However, all the coefficients in the sum are 0:

$$\langle T(v_k), v_j \rangle = \langle v_k, \overline{\lambda_j} v_j = \overline{\lambda_j} \langle v_k, v_j \rangle = 0$$

Thus, $T(v_k) = \langle T(v_k), v_k \rangle v_k$, so $T(v_k) = \lambda_k v_k$.

This can be proven up to n, so $\{v_1, \ldots, v_n\}$ are all eigenvectors. ■

This means that all self-adjoint (Hermitian) matrices—complex matrices such that $A = A^*$—can be diagonalized.

Definition 7.13. Let $T \in \mathcal{L}(V)$, where V is finite dimensional and an inner product space. We say that T is self-adjoint or Hermitian if $T^* = T$.

Lemma 7.3. Let T be a self-adjoint linear operator on a finite-dimensional inner product space V. then

1. Every eigenvalue of T is real;
2. If V is a real inner product space, then the characteristic polynomial of T splits.

Proof. If x is an eigenvector for T, then it is also an eigenvector of T^*.

$$\lambda x = Tx = T^* x = \overline{\lambda} x$$

Therefore, $\lambda = \overline{\lambda}$, so it is real.

Next, see $[T]$ as a matrix with complex entries. Therefore, T splits over $\mathbb{C}$, but because it is self-adjoint, its eigenvalues are real. Therefore, the characteristic polynomial splits over $\mathbb{R}$. ■

Now, we turn to one of the most important theorems covered in these notes: [MARK FOR BOXING]

Theorem 7.4. Let $T \in \mathcal{L}(V)$ where V is a finite-dimensional real inner product space. Then T is self-adjoint if and only if there exists an orthonormal basis β consisting of eigenvectors of T.

Proof. Suppose there exists such a β. Then $[T]_\beta$ is diagonal, but $[T^*]_\beta$ is also diagonal; because the eigenvalues are real, so $T^* = T$.

Next, by the previous lemma, if T is self-adoint, then its characteristic polynomial splits. By Schur's theorem, $[T]_\beta$ is upper triangular, and by assumption so is $[T^*]_\beta$. On the other hand, $[T^*]_\beta = [(T^t)^*]_\beta$ because it is self-adjoint and real. Therefore, this can only occur if $[T]_\beta$ is diagonal, as desired. ■

All the theorems we have seen about adjoints and diagonalizing/triangularizing are true for matrices. However, we don't need to define an inner product, because we can technically define them in terms of transposition and conjugation.

This arises from the fact that there is an implicit inner product for matrices, wich is the canonical one:

$$\langle x, y \rangle = \sum x_i \overline{y}_i$$

In particular, you would find

$$\langle x, Ay \rangle = \langle A^* x, y \rangle$$

If x, y are column vectors in $\mathbb{F}^n$, then their inner product is just

$$x^t \overline{y}$$

Then

$$\langle x, Ay \rangle = x^t \overline{Ay} = x^t \overline{A}\overline{y} = \langle (A^* x)^t, y \rangle$$

7.5 Unitary and Orthogonal Operators

We have seen that normal operators and matrices on a complex inner product space are diagonalizable on an orthonormal basis.

Definition 7.14. Let $T \in \mathcal{L}(V)$, where V is finite-dimensional. If

$$\|T(x)\| = \|x\|$$

for all $x \in V$, then we call T a unitary operator if $\mathbb{F} = \mathbb{C}$, and an orthogonal operator if $\mathbb{F} = \mathbb{R}$.

Note that in the infinite-dimensional case, an operator satisfying these conditions is generally called an isometry.

Theorem 7.5. The following conditions are equivalent:

1. $TT^* = T^*T = I$ $(T^* = T^{-1})$;
2. $\langle T(x), T(y)\rangle = \langle x, y\rangle$ for all $x, y \in V$;
3. If β is an orthonormal basis for V, then $T(\beta)$ is an orthonormal basis for V;
4. There exists an orthonormal basis β for V such that $T(\beta)$ is an orthonormal basis for V.
5. $\|T(x)\| = \|x\|$ for all $x \in V$.

Remark. T is orthogonal if and only if its eigenvalues satisfy $|\lambda_i| \leq 1$

Definition 7.15. A matrix A is orthogonal if $A^tA = AA^t = I$ (real case), and unitary $A^*A = AA^* = I$ (complex case).

A matrix is unitary/orthogonal if its column vectors form an orthonormal basis for the canonical inner product. This proposition can be checked computationally.

We learned that self-adjoint or Hermitian linear maps can be diagonalized in an orthonormal basis. For a matrix, that means

$$A = P^{-1}DP$$

Where D is a diagonal matrix; this gives you the coordinates of the new matrix in terms of the old ones (or vice versa). Here, the new basis is orthonormal, whereas the old basis was the canonical one. Therefore, P must be an orthogonal/unitary matrix. Hence, $P^*P = I$ and $P^{-1} = P^*$. Thus, A can be written as

$$A = P^*DP$$

This leads to the following definition:

Definition 7.16. We say that A and B are unitary equivalent matrices if there exists a unitary/orthogonal transformation P such that

$$A = P^*BP$$

Therefore, we see that normal/self-adjoing matrices are unitarily equivalent to diagonal matrices. This is actually an equivalence relation:

Theorem 7.6. *A is a normal/self-adjoint operator if and only if A is unitarily equivalent to a diagonal matrix.*

7.6 Bilinear and Quadratic Forms

For now, let V be an inner product space over the real numbers.

Definition 7.17. A function $H : V \times V \to \mathbb{F}$ is called bilinear if it is linear with respect to each variable

An example of this is the scalar product over the real numbers. However, note that a bilinear form does not always constitute an inner product.

How do we determine this bilinear form? Given a basis $\beta = \{v_1, \ldots, v_n\}$, then:

$$x = \sum x_i v_i$$

$$y = \sum y_j v_j$$

$$H(x, y) = \sum_i \sum_j x_i y_j H(v_i, v_j)$$

Therefore the data of $H(v_i, v_j)$ suffices to determine the bilinear form. This data has two indecies, so we put it in an $n \times n$ matrix A where

$$A_i j = H(v_i, v_j)$$

It can be seen that

$$H(x, y) = x^t A y$$

This is an important formula, because it connects a bilinear form to the matrix that represents it.

Warning. Note—this is not the same as a matrix representing a linear map. Note that the coordinates of the matrix A are no longer the images of the basis vectors after the transformation.

We are now free to state some properties about bilinear forms:

Theorem 7.14. The following are properties for bilinear forms:

1. The set of all bilinear forms, denoted $\mathcal{B}(V)$ on a vector space V is itself a vector space ($aH_1 + H_2$ is also a bilinear form)
2. The aforementioned vector space of bilinear forms is of dimension n^2, because $\mathcal{B}(V)$ is isomorphic to a matrix of size $n \times n$.

Now we can ask: how can we change the basis of a bilinear form? It turns out that there's a formula:

Theorem 7.15. If you have two bases β and γ, and $Q = [I]_\beta^\gamma$, then

$$\psi_\gamma(H) = Q^t \psi_\beta(H) Q$$

Where ψ is the representation of the bilinear form in a specific basis. Here the relationship between the matrices is called congruence. Note that Q must be invertible.

Definition 7.18. We say that a bilinear form is symmetric if

$$H(x, y) = H(y, x)$$

It follows that the matrix that represents H is also symmetric.

7.6.1 Sesquilinear Forms

In the complex case, we define what are called sesquilinear forms.

Definition 7.19. A sesquilinear form is a map $H : V \times V \to \mathbb{C}$ such that it satisfies the definition of a bilinear form except that

$$H(x, \lambda y) = \overline{\lambda} H(x, y)$$

and

$$H(x, y) = \overline{H(y, x)}$$

The matrix representing H is therefore self-adjoint.

7.6.2 Quadratic Forms

Definition 7.20. A function $K : V \to \mathbb{F}$ is a quadratic form if

$$K(x) = H(x, x)$$

for some symmetric bilinear form H.

In coordinates, recall how the bilinear form was determined on the basis vectors. Similarly, quadratic forms are written like:

$$K(x) = \sum_i \sum_j x_i x_j H(v_i, v_j)$$

Note that familiar conic sections are of the form

$$K(x) = c$$

Our goal is to write quadratic forms into the simplest way; that is, we prefer to have a sum of squares.

Another thing to note is that if K is known, we can always recover H on which it is based. In particular:

$$H(x, y) = \frac{1}{2}[K(x + y) - K(x) - K(y)]$$

Which comes from how quadratic functions are expanded and reduced.

Now, reducing H to a sum of squares corresponds to finding a basis β for which its matrix is diagonal. This is equivalent to a quadratic form.

If the bilinear form is symmetric (if not, we do not know much about it), then any matrix representing it is also symmetric. But going back to what we know about matrices, we know that its matrix can be diagonalized in an orthonormal basis, so it is congruent to a diagonal matrix.

We can therefore conclude that up to a change of basis, the matrix representing H is diagonal. This is called diagonalizing the bilinear form.

Theorem 7.16. Every symmetric bilinear form is diagonalizable.

Moreover, if $H(v_i, v_i) > 0$, then we can set

$$w_i = \frac{v_i}{H(v_i, v_i)}$$

Then

$$H(w_i, w_i) = 1.$$

If $H(v_i, v_i) < 0$,

$$w_i = \frac{v_i}{-H(v_i, v_i)}$$

So then

$$H(w_i, w_i) = -1$$

Thus, any quadratic form can be written as a matrix with 1's,-1's, and 0's along the main diagonal and 0 everywhere else. This means that in the proper coordinates, they can be written as

$$x_1^2 \pm, \ldots, \pm x_k^2$$

However, the number of $+$'s and $-$'s are always invariant. This property of quadratic forms is known as the signature. If and only if they're all 1's, then

$$H(\cdot, \cdot)$$

defines an inner product.

As a result, conics can always be written in the form

$$x_1^2 \pm, \ldots, \pm x_k^2 = c$$

8 Jordan Canonical Forms

When we cannot diagonalize a linear map, we still seek to put it in its simplest form (after changing the basis.) The simplest possible form is the Jordan Canonical Form, which is possible if the characteristic polynomial splits.

It is a matrix in block diagonal form:

$$\begin{bmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_k \end{bmatrix}$$

Where each J_i is known as a Jordan Block, and is either of the form $[\lambda]$ or

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

All λ are eigenvalues of T/A; in particular, a diagonal matrix is thus in Jordan block form. Jordan block matries are upper triangular, so their determinats may easily be computed.

Notice how the form can be written as

$$\lambda I + M$$

Where M is a matrix with 1's shifted to the right of the diagonal. It is worth noting that M is a nilpotent matrix; M^n is the zero map where n is the dimension of the matrix.

Therefore, the only eigenvalue of a nilpotent matrix is zero (otherwise, if there were nonzero entries on the diagonal, in multiplication they would simply multiply with themselves).

Definition 8.1 (Generalized Eigenvectors)**.** For a linear map T, a vector x is called a generalized eigenvector of T for λ if there exists a positive integer p such that:

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

Note that this definition includes regular eigenvectors in the case $p = 1$. Moreover, λ is an eigenvalue whose eigenvector is

$$(T - \lambda I)^{p-1}(x)$$

Definition 8.2 (Generalized Eigenspace)**.** We then let

$$K(\lambda, T) = \{x \in V : (T - \lambda)^p(x) = 0, p \in \mathbb{Z}\}$$

for some positive integer p.

Theorem 8.1. The generalized eigenspace has the following properties:

1. $K(\lambda, T)$ is T-invariant, and contains $E(\lambda, T)$.
2. For any $\mu \neq \lambda$, the restriction of $T - \mu I$ to $K(\lambda, T)$ is one-to-one.

Theorem 8.2. Let $T \in \mathcal{L}(V)$, where V is finite dimensional, such that the characteristic polynomial of T splits. Suppose λ is an eigenvalue with multiplicity m. Then

1. $\dim(K(\lambda, T)) \leq m$;
2. $K(\lambda, T) = \ker(T - \lambda I)^m$.

Theorem 8.3. Let $T \in \mathcal{L}(V)$, such that the characteristic polynomial of T splits. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then for every $x \in V$, there exist generalized eigenvectors $v_i \in K(\lambda_i, T)$ such that

$$x = v_1 + \cdots + v_k$$

Theorem 8.1. Let $T \in \mathcal{L}(V)$, such that the characteristic polynomial of T splits. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T with multiplicity m_i. Let β_i be a basis for $K(\lambda_i, T)$. Then

1. $\beta_i \cap \beta_j = \varnothing$ for $i \neq j$.
2. $\beta = \beta_1 \cup \cdots \cup \beta_k$ is an ordered basis for V.
3. $\dim K(\lambda_i, T) = m_i$.

Corollary 8.4. As a result, we observe that T is diagonalizable if and only if

$$E(\lambda, T) = K(\lambda, T)$$

for every eigenvalue λ.

[THIS SECTION UNDER CONSTRUCTION]

8.1 The Minimal Polynomial

Recall how when we discussed linear maps, we considered linear maps in polynomials; that is,

$$p(T) = a_0 I + a_1 T + \cdots + a_m T^m$$

We have seen that if f_T is the characteristic polynomial of T, then T is a solution to its own characteristic polynomial by the Cayley-Hamilton theorem. Note that this polynomial is of degree $n = \dim V$.

The question is, are there any other polynomials $g(T)$ such that $g(T) = 0$? In short, yes:

Remark. The polynomial

$$g(t) = (h \circ f_T)(t)$$

also admits T as a solution.

Definition 8.3. A polynomial p is called the minimal polynomial of T if $p(t)$ is a monic polynomial of least positive degree for which

$$p(T) = 0$$

Note that monic simply means the leading coefficient is 1. We know that due to the Cayley-Hamilton theorem, there exists an upper bound to this degree, namely n, so a minimal polynomial does indeed exist.

Corollary 8.5. If $g(T)$ is a polynomial that "cancels" T, then $p(t)$ must divide $g(t)$. That is,

$$g(t) = q(t)p(t)$$

Proof. We know that the degree of g must be greater than or equal to the degree of p, otherwise we have a contradiction in our assumptions.

Therefore, we can do a euclidean division, to get g in the form

$$g(t) = q(t)p(t) + r(t)$$

We see that the degree of r is strictly less than the degree of p. Inputting T into the equation:

$$g(T) = q(T)p(T) + r(T)$$

Because $g(T)$ and $p(T)$ are both zero, we conclude that $r(T)$ is zero. But then r would be a polynomial of smaller degree that cancels T, so we arrice at a contradiction, unless r is 0. Therefore, there is no remainder, so

$$g(T) = q(T)p(T)$$

■

Corollary 8.6. The minimal polynomial divides the characteristic polynomial.

If T is a linear operator, then its minimal polynomial is the same as the minimal polynomial of its matrix representation.

Theorem 8.7. Let $T \in \mathcal{L}(V)$ with minimal polynomial $p(T)$. A scalar λ is an eigenvalue of T if and only if

$$p(\lambda) = 0$$

Hence $p(t)$ and $f_T(t)$ have the same zeros.

Proof. Suppose λ is a zero of $p(t)$. We know that

$$f_T(t) = q(t)p(t)$$

Therefore λ is a zero of the characteristic polynomial, so it is an eigenvalue of T.

Conversely, suppose λ is an eigenvalue of T. Then we know that $f(\lambda) = 0$, so let x be an eigenvector corresponding to λ:

$$p(T)(x) = p(\lambda)x = 0$$

Because we know x is not zero, $p(\lambda)$ must be zero and hence it is a zero for the minimal polynomial. ■

Corollary 8.8. With the same assumptions, if f_T splits, then

$$f_T(t) = (\lambda_1 - t)^{m_1} \cdots (\lambda_k - t)^{m_k}$$

with λ_i being the distinct eigenvalues of T, and m_i being their multiplicities, then

$$p(t) = (t - \lambda_1)^{n_1} \cdots (t - \lambda_k)^{n_k}$$

where $1 \leq n_i \leq m_i$.

Theorem 8.9. Let $T \in \mathcal{L}(V)$, such that V is an n-dimensional T-cyclic subspace of itself. Then the characteristic polynomial $f_T(t)$ and the minimal polynomial $p(t)$ have the same degree, and hence

$$f_T(t) = (-1)^n p(t)$$

Proof. Recall that by the definition of T-cyclic subspace, there is an x such that

$$\beta = \{x, T(x), \dots, T^{n-1}(x)\}$$

is a basis for V. Let

$$g(t) = a_0 + \cdots + a_k t^k$$

be a polynomial of degree $k < n$. Then $a_k \neq 0$ and

$$g(T)(x) = a_0 x + a_1 T(x) + \cdots + a_k T^k(x)$$

However, these vectors $x, \dots, T^k(x)$ are linearly independent, so $g(T)(x)$ cannot be zero unless a_i are all equal to zero. Therefore, g cannot be the minimal polynomial. Therefore, the minimal polynomial cannot have degree less than n, so it is of degree n. So the characteristic polynomial is the same as the minimal polynomial, up to a multiplicative constant. ■

What these theorems build up to is

Theorem 8.2. if $T \in \mathcal{L}(V)$, and V is finite-dimensional, then T is diagonalizable if and only if the minimal polynomial of T is of the form

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

where λ_i are the distinct eigenvalues of T.

Proof. We will prove this with induction on the dimension. One direction is easy;

Suppose T is diagonalizable, with distinct eigenvalues $\lambda_1 \dots, \lambda_k$. Set

$$f(t) = (t - \lambda_1) \cdots (t - \lambda_k)$$

We can check that $f(T)$ is zero, because there exists a basis of eigenvectors for V, denoted v_i's. We see that

$$f(T)(v_i) = (T - \lambda_1) \cdots (T - \lambda_i)(v_i)$$

The order can be changed because they commute. One of them cancels v_i, so the entire expression $f(T)(v_i) = 0$. Therefore, $f(T)$ maps the entire basis to zero; hence, $f(T) = 0$.

So the minimal polynomial p divides f, and has the same roots as f, because all the eigenvalues have to be zeroes of p. Therefore, $p = f$.

Conversely, assume $f(T) = 0$; we need to show that T is diagonalizable.

By induction, let's assume that T is diagonalizable. The case for $\dim V = n = 1$ follows immediately. Now assume it has been proven up to $n - 1$. We will try to work with the eigenspace, and show that it has a complement space that is T-invariant. Let λ_k be the last eigenvalue in the list of eigenvalues, and let's set $E(\lambda_k, T)$, which is T-invariant. Let's also set $W = \Im(T - \lambda_k I)$, which is also T-invariant. By the dimension theorem,

$$\dim W = n - \dim E(\lambda_k, T) \leq n - 1$$

because the dimension of the eigenspace must be at least one. Furthermore, we need to check that

$$E(\lambda_k, T) \cap W = \{0\}$$

This is because λ_k is not an eigenvalue of $T|_W$. If not, then there exists an $x \in V$ such that

■