

MATH 3406 Final Portfolio

Daniel Öman

December 14, 2023

Contents

1	Introduction	1
2	Linear Maps	1
2.1	An Introduction to Linear Maps	1
2.2	The Fundamental Theorem of Linear Maps	2
2.3	Matrices of Linear Maps	5
2.4	Discussion	6
3	Orthogonal Complements and Minimization Problems	7
3.1	Orthogonal Complements	7
3.1.1	Direct Sum of a Subspace and Its Orthogonal Complement	7
3.2	Minimization Problems and How They Relate to Orthogonal Complements	9
3.2.1	Minimization with the Orthogonal Projection	9
3.3	Discussion	12
4	Eigenvectors and Eigenvalues	12
4.1	Basics of Eigenvalues and Eigenvectors	13
4.2	Upper Triangular and Diagonal Matrices	14
4.2.1	Upper Triangular Matrices of Operators	15
4.2.2	Diagonal Matrices of Operators	16
4.3	Discussion	17
5	Markov Chains	17
5.1	A Practical Introduction to Markov Chains	18
5.2	Steady States	19
5.2.1	Steady State Vectors	19
5.2.2	The Matrix of a Transition Operator	20
5.2.3	The Perron-Frobenius Theorem and the Steady State	23
5.3	Discussion	24
A	Additional Proofs and Results	26
A.1	Proof that Differentiation on Polynomials is a Linear Map	26
A.2	Inner Product from Section 3, Problem 3	26
A.3	Eigenvalues of Normal Operators	27

1 Introduction

MATH 3406: A Second Course in Linear Algebra at Georgia Tech has taught me linear algebra from a fresh new perspective. In this portfolio, I will explore linear maps, orthogonal complements and minimization problems, and eigenvectors and eigenvalues. In these sections, I will compile a collection of definitions, theorems, proofs, and problems, drawn from Sheldon Axler's textbook (and this course's book) *Linear Algebra Done Right* [1]. To tie the portfolio together, I delve deep into Markov Chains, a fascinating application of so many different parts of linear algebra.

2 Linear Maps

Linear maps are a core concept in linear algebra. The field of study is called “linear” algebra, after all! In this section, I will explore a variety of definitions, proofs, and problems related to linear maps and their nuances.

2.1 An Introduction to Linear Maps

I will begin by introducing linear maps through a definition.

Definition 1. A linear map is a function that takes vectors in some vector space V to another vector space W , over some scalar field \mathbb{F} , that satisfies additivity and homogeneity: Suppose T is a linear map from V to W . Then, T satisfies

- Additivity: $T(u + v) = Tu + Tv$, for all $u, v \in V$
- Homogeneity: $T(\lambda v) = \lambda Tv$, for all $v \in V$, $\lambda \in \mathbb{F}$.

The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$. If $W = V$, then T is called a linear operator and the shorthand notation $\mathcal{L}(V, V) = \mathcal{L}(V)$ is used. That is, if T is an operator on V , then $T \in \mathcal{L}(V)$.

An important fact of linear maps in $\mathcal{L}(V, W)$ is that they take $0 \in V$ to $0 \in W$.

Proof. Suppose $T \in \mathcal{L}(V, W)$. Then, by the additivity requirement, $T(0) = T(0 + 0) = T(0) + T(0)$. Now, since $T(0) \in W$, and W is a vector space, then there exists some additive inverse $w \in W$ such that $T(0) + w = 0$. So, we have that

$$T(0) = T(0) + T(0) \Rightarrow T(0) + w = T(0) + T(0) + w \Rightarrow 0 = T(0),$$

as desired. □

With this knowledge, we can solve our first example problem.

Example Problem 1: (3.A #4) Suppose $T \in \mathcal{L}(V, W)$ and v_1, \dots, v_m is list of vectors in V such that Tv_1, \dots, Tv_m is a linearly independent list in W . Prove that v_1, \dots, v_m are linearly independent.

Solution: I will proceed with a direct proof.

Proof. Suppose we have $c_1v_1 + \dots + c_mv_m = 0$ for some $c_1, \dots, c_m \in \mathbb{F}$. Then, applying T to both sides gives

$$T(c_1v_1 + \dots + c_mv_m) = T(0).$$

By the proof above, $T(0) = 0$, so

$$T(c_1v_1 + \cdots + c_mv_m) = 0.$$

By additivity,

$$T(c_1v_1) + \cdots + T(c_mv_m) = 0,$$

and by homogeneity,

$$c_1Tv_1 + \cdots + c_mTv_m = 0.$$

However, we know that Tv_1, \dots, Tv_m are linearly independent, and they are linearly independent if and only if $c_1 = \cdots = c_m = 0$. Therefore, this implies that $c_1v_1 + \cdots + c_mv_m = 0$ if and only if $c_1 = \cdots = c_m = 0$, so v_1, \dots, v_m are linearly independent. \square

This problem neatly shows the interaction between vector spaces and linear maps. However, we can note that the opposite direction is not always true. That is, if v_1, \dots, v_m are linearly independent, then Tv_1, \dots, Tv_m are not always linearly independent. For example, consider $T = 0$. Clearly, $Tv_1 = \cdots = Tv_m = 0$, so they are trivially not linearly independent, even though v_1, \dots, v_m are!

2.2 The Fundamental Theorem of Linear Maps

Two key definitions in the study of linear maps are *injectivity* and *surjectivity*. The next two definitions are not limited to linear maps, but I will introduce them in this context, as it is most relevant here.

Definition 2. A linear map $T \in \mathcal{L}(V, W)$ is injective if $Tu = Tv$ implies that $u = v$.

Definition 3. A linear map $T \in \mathcal{L}(V, W)$ is surjective if $\{Tv : v \in V\} = W$.

The second definition introduces the notion of the *range* of a linear map. The range of a linear map $T \in \mathcal{L}(V, W)$ is the subset of W defined as $\text{range } T = \{Tv : v \in V\}$. The range of a linear map has a close sibling, the null space. The null space of $T \in \mathcal{L}(V, W)$ is defined by $\text{null } T = \{v \in V : Tv = 0\}$, which is a subset of V .

An important fact is that the null space is a subspace of V , and the range is a subspace of W .

Proof. Suppose $T \in \mathcal{L}(V, W)$. I will first prove that the null space is a subspace of V . Suppose $u, v \in \text{null } T$. Then, $Tu = 0$ and $Tv = 0$. Now, $T(u + v) = Tu + Tv$ by linearity, and $Tu + Tv = 0 + 0 = 0$, by definition of u and v . So, $T(u + v) = 0$, and so $u + v \in \text{null } T$. Now, take any $\lambda \in \mathbb{F}$. Then, $T(\lambda v) = \lambda Tv$ by homogeneity, and $\lambda Tv = \lambda 0 = 0$ by definition of v . So, $T(\lambda v) = 0$, and therefore $\lambda v \in \text{null } T$. Finally, $0 \in \text{null } T$ because $T(0) = 0$ for any linear map. Therefore, the null space is a subspace of V .

I will now prove that the range is a subspace of W . Suppose $w, x \in \text{range } T$. Then, by the definition of the range, there exist some $u, v \in V$ such that $Tu = w$ and $Tv = x$. Now, $w + x = Tu + Tv = T(u + v)$ by linearity. Then, set $u' = u + v$, and we have that $w + x = Tu'$. So, $w + x \in \text{range } T$, because there is some vector in V that maps to $w + x$ in W . Next, take any $\lambda \in \mathbb{F}$. If $w \in \text{range } T$, then there exists $v \in V$ such that $Tv = w$. Multiplying both sides by λ we get $\lambda Tv = \lambda w$, and by homogeneity, $T(\lambda v) = \lambda w$. Set $v' = \lambda v$, so $Tv' = \lambda w$. Therefore, $\lambda w \in \text{range } T$. Finally $0 \in \text{range } T$ because there exists some $v \in V$ such that

$Tv = 0$. This v is at least $v = 0$, although there may be more, depending on the linear map. Therefore, the range is a subspace of W . \square

With these key definitions in place, I now want to introduce the *Fundamental Theorem of Linear Maps*, a very “grand-sounding” theorem in linear algebra and the study of linear maps:

Theorem 1. If we have some finite-dimensional vector space V , another vector space W , and a linear map $T \in \mathcal{L}(V, W)$, then

$$\dim V = \dim \text{null } T + \dim \text{range } T.$$

This is a very important theorem, and it is used in many dimensionality arguments for problems and proofs in linear algebra. Therefore, I will prove it in my own words.

Proof. Suppose V , W , and T are defined as in the theorem’s statement. Then, let u_1, \dots, u_m be a basis of $\text{null } T$. Since $\text{null } T$ is a subspace of V , I can add n new vectors to extend the linearly independent list to form a basis of V , resulting in the basis of V given by $u_1, \dots, u_m, v_1, \dots, v_n$. Now, I have that $\dim V = m + n$, and $\dim \text{null } T = m$, so $\dim V = \dim \text{null } T + m$. All that is left now is showing that $n = \dim \text{range } T$.

To proceed, I pick some arbitrary $v \in V$, and write it in terms of the basis of V I gave. For this, consider $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$. Then, $v = a_1u_1 + \dots + a_mu_m + b_1v_1 + \dots + b_nv_n$. Next, I apply T to both sides to get

$$Tv = T(a_1u_1 + \dots + a_mu_m + b_1v_1 + \dots + b_nv_n).$$

By linearity and homogeneity, this is

$$Tv = a_1Tu_1 + \dots + a_mTu_m + b_1Tv_1 + \dots + b_nTv_n.$$

Since $u_1, \dots, u_m \in \text{null } T$, by definition, $Tu_j = 0$ for $j = 1, \dots, m$. So, we get that

$$Tv = b_1Tv_1 + \dots + b_nTv_n.$$

This linear combination demonstrates that Tv_1, \dots, Tv_n spans $\text{range } T$, because we wrote any arbitrary $v \in V$ in terms of a linear combination of Tv_1, \dots, Tv_n . What remains in the proof is to show that Tv_1, \dots, Tv_n are linearly independent.

For this, suppose we have $c_1, \dots, c_n \in \mathbb{F}$ such that $c_1Tv_1 + \dots + c_nTv_n = 0$. By linearity and homogeneity, this becomes

$$T(c_1v_1 + \dots + c_nv_n) = 0,$$

and therefore, $c_1v_1 + \dots + c_nv_n \in \text{null } T$, by definition. So, we can introduce a new set of coefficients, $d_1, \dots, d_m \in \mathbb{F}$ and write $c_1v_1 + \dots + c_nv_n$ in terms of our previously defined basis of $\text{null } T$:

$$c_1v_1 + \dots + c_nv_n = d_1u_1 + \dots + d_mu_m.$$

This implies that

$$c_1v_1 + \dots + c_nv_n - d_1u_1 - \dots - d_mu_m = 0,$$

and since $v_1, \dots, v_n, u_1, \dots, u_m$ form a basis of V , they are linearly independent, and they are linearly independent if and only if $c_1 = \dots = c_n = \dots d_1 = \dots d_m = 0$. This implies that

$$c_1Tv_1 + \dots + c_nTv_n = 0$$

if and only if $c_1 = \cdots = c_n = 0$, so Tv_1, \dots, Tv_n are linearly independent. Therefore, they not only span range T , but form a basis of range T , and thus, $\dim \text{range } T = n$.

We now have that $\dim V = \dim \text{null } T + \dim \text{range } T$, as desired. \square

This proof is the ultimate dimensionality argument in action in a problem solving context. The next two problems also argue using dimensionality.

Example Problem 2: (6.B #6) Prove that there does not exist a linear map $T \in \mathcal{L}(\mathbb{R}^5, \mathbb{R}^5)$ such that $\text{range } T = \text{null } T$.

Solution: I will proceed using a proof by contradiction.

Proof. Suppose that there does exist such a T . Then, under the fundamental theorem of linear maps, $\dim \mathbb{R}^5 = 5 = \dim \text{null } T + \dim \text{range } T$. If $\text{range } T = \text{null } T$, then $\dim \text{range } T = \dim \text{null } T$, because I could create one basis of some integer length for both vector spaces. This implies that $5 = \dim \text{range } T + \dim \text{range } T$, so $5 = 2 \dim \text{range } T$. Thus, this requires that $\dim \text{range } T = 5/2$, which is not an integer. This is a contradiction, because the dimension of a vector space must be an integer. Therefore, there does not exist such a T , completing our proof. \square

In fact, this problem can be extended to any $T \in \mathcal{L}(V, W)$ such that $\dim V$ is odd. We can set $\dim V = 2k + 1$ for any integer $k \geq 0$, and then our proof will lead to the contradiction that $\dim \text{range } T = (2k + 1)/2$, implying that $\dim \text{range } T = k + 1/2$, which is not an integer.

This next problem is a bit more nuanced.

Example Problem 3: (6.B #17) Suppose V and W are both finite-dimensional. Prove that there exists an injective linear map from V to W if and only if $\dim V \leq \dim W$.

Solution: To begin this problem, it is important to first note that $T \in \mathcal{L}(V, W)$ is injective if and only if $\text{null } T = \{0\}$. I will first prove this detail.

Proof. (\rightarrow) First, suppose T is injective. Since $\text{null } T$ is a vector space, $\{0\} \subset \text{null } T$. Now, pick any $v \in \text{null } T$. By definition of the null space, $Tv = 0 = T(0)$. Since T is injective, then $v = 0$ is the only vector which satisfies this condition, so $\text{null } T \subseteq \{0\}$. Thus, $\text{null } T = \{0\}$.

(\leftarrow) If $\text{null } T = \{0\}$, then pick $u, v \in V$, such that $Tu = Tv$. Then, $0 = Tu - Tv = T(u - v)$, so $T(u - v) = 0$. Therefore, $u - v \in \text{null } T = \{0\}$, so $u - v = 0$. Therefore, $u = v$, and thus T is injective. \square

Now that we have this fact proven, I can proceed with the problem.

Proof. (\rightarrow) If T is injective, then $\text{null } T = \{0\}$, and so $\dim \text{null } T = 0$. So, $\dim V = \dim \text{range } T$, by the fundamental theorem of linear maps. The range is a subspace of W , so $\dim \text{range } T \leq \dim W$, and thus $\dim V \leq \dim W$.

(\leftarrow) Suppose that $\dim V \leq \dim W$. Let v_1, \dots, v_n be a basis for V and w_1, \dots, w_m be a basis of W . Then, there exists a linear map T such that $Tv_1 = w_1, \dots, Tv_n = w_n$. This T can be defined by $T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$ for arbitrary $c_1, \dots, c_n \in \mathbb{R}$.

\mathbb{F} (see the proposition 3.5 in *Linear Algebra Done Right* for a full proof of the existence of this T). Next, pick a $c_1v_1 + \cdots + c_nv_n \in \text{null } T$. Then, $T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n = 0$. Each w_1, \dots, w_n are linearly independent since they are part of a basis for W , so it must be the case that $c_1 = \cdots = c_n = 0$. So, $\text{null } T \subset \{0\}$, as we picked any arbitrary vector. However, we know that $\{0\} \subseteq \text{null } T$, because the null space is a vector space. This implies that $\{0\} = \text{null } T$, which means that this T is injective, by the previous proof. Therefore, there does exist such a T , completing both directions of the proof. \square

Both of these problems use dimensionality of finite dimensional subspaces as their key argument, a powerful tool in solving problems in linear algebra. We will see this tool come up in the later sections of this portfolio, especially in section 4.

2.3 Matrices of Linear Maps

When most people discuss linear algebra, one of the first concepts that may come to mind are matrices, the two-dimensional grids of numbers. In linear algebra, a matrix is simply a representation of a linear map $T \in \mathcal{L}(V, W)$ under particular choices of bases for V and W .

Definition 4. Suppose $T \in \mathcal{L}(V, W)$, the vectors v_1, \dots, v_n form a basis of V , and the vectors w_1, \dots, w_m form a basis of W . Then, the matrix of T , denoted $\mathcal{M}(T)$ under the particular choice of basis, is defined as the m -by- n matrix whose entry at row i , column j , are given by $c_{i,j} \in \mathbb{F}$, for $i = 1, \dots, m$, $j = 1, \dots, n$, and each $c_{i,j}$ is defined by

$$Tv_j = c_{1,j}w_1 + \cdots + c_{i,j}w_i + \cdots + c_{m,j}w_m.$$

The entry at row i , column j describes what happens to the i th basis vector of W , w_i , when applying T to the j th basis vector of V , v_j .

Using matrices to represent linear maps have a vast amount of practical applications, as it is easier to encode a transformation as a matrix in a computer than it is to represent a linear map on its own. In section 5, I explore a real-world application involving the matrices of a particular kind of linear map, the transition operator of a Markov Chain.

I will now work through two problems tying together many concepts from linear maps involving matrices.

Example Problem 4: (3.C #1) Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Show that with respect to each choice of bases of V and W , the matrix of T has at least $\dim \text{range } T$ nonzero entries.

Solution: For this problem I proceed with a direct proof.

Proof. First, pick v_1, \dots, v_n to be an arbitrary basis of V and w_1, \dots, w_m to be an arbitrary basis of W . Under this choice of basis, we have that $\mathcal{M}(T)$ is the matrix of T . Next, without loss of generality, suppose that out of the n basis vectors of V , the first k of them map to zero. That is, $Tv_1 = \cdots = Tv_k = 0$. If this is the case, then by the definition of $\mathcal{M}(T)$, the first k columns will be entirely zero. Under the same assumption, we therefore have $n - k$ remaining vectors such that each $Tv_j \neq 0$, for $j = k+1, k+2, \dots, n$. By the definition of $\mathcal{M}(T)$, each Tv_j will correspond to a column

of $\mathcal{M}(T)$ which will have at least one nonzero entry, since the vector is not mapped to zero. One of the m coefficients in the expression of Tv_j in terms of w_1, \dots, w_m will be nonzero, so at least one entry in the remaining $n - k$ columns will be nonzero. We can also note that $\text{range } T = \text{span}(v_{k+1}, \dots, v_n)$, because they can form any non-zero vector in W that T maps to. Note that not all of v_{k+1}, \dots, v_n are linearly independent, and so we must pick at least $\dim \text{range } T$ of them to form a linearly independent spanning set of $\text{range } T$, i.e., a basis of T . So, there will be at least $\dim \text{range } T$ vectors that correspond to a column with at least one nonzero entry, and thus, there will be at least $\dim \text{range } T$ nonzero entries in $\mathcal{M}(T)$. \square

The next problem requires more computation than reasoning, but it is helpful to illustrate matrices and an example of a linear map.

Example Problem 5: (3.C #2) Suppose $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ is the differentiation map defined by $Dp = p'$ for all $p \in \mathcal{P}_3(\mathbb{R})$. Find a basis of \mathcal{P}_3 and \mathcal{P}_2 such that the matrix of D is

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

(see appendix A.1 for a proof of why D is a linear map!)

Solution: For the sake of this problem we can pick any basis of $\mathcal{P}_2(\mathbb{R})$ to start, and work towards finding a basis of $\mathcal{P}_3(\mathbb{R})$ which has the corresponding matrix. Let's pick the standard basis of $\mathcal{P}_2(\mathbb{R})$, $1, x, x^2$. Now, suppose p_1, p_2, p_3, p_4 form a basis of $\mathcal{P}_3(\mathbb{R})$. We can write this basis in terms of the standard basis $1, x, x^2, x^3$ as:

$$p_1 = a_0 + a_1x + a_2x^2 + a_3x^3$$

$$p_2 = b_0 + b_1x + b_2x^2 + b_3x^3$$

$$p_3 = c_0 + c_1x + c_2x^2 + c_3x^3$$

$$p_4 = d_0 + d_1x + d_2x^2 + d_3x^3$$

We need to choose $a_0, \dots, a_3, b_0, \dots, b_3, c_0, \dots, c_3, d_0, \dots, d_3 \in \mathbb{R}$ such that $Dp_1 = 1$, $Dp_2 = x$, $Dp_3 = x^2$, and $Dp_4 = 0$. We can easily set $d_0 = t$ where $t \in \mathbb{R}$ is an arbitrary nonzero constant, and $d_1 = d_2 = d_3 = 0$ to get $p_4 = t$. It is important that $t \neq 0$ because we need that p_4 be a basis vector, which cannot be zero. We see that $Dp_4 = 0$, so the fourth column will be entirely zeros, as desired. Now, $Dp_3 = c_1 + 2c_2x + 3c_3x^2$, so we need to set $c_1 = c_2 = 0$ and $c_3 = 1/3$. So, $p_3 = (1/3)x^3$. Next, $Dp_2 = b_1 + 2b_2x + 3b_3x^2$, so we must set $b_2 = 1/2$ and $b_1 = b_3 = 0$, giving us $p_2 = (1/2)x^2$. Finally, we see that $Dp_1 = a_1 + a_2x + a_3x^2$, so we can easily set $a_1 = 1$ and $a_2 = a_3 = 0$ to get $p_1 = (1/2)x$. So, our final basis of $\mathcal{P}_3(\mathbb{R})$ is $t, x, (1/2)x^2, (1/3)x^3$, where $t \neq 0$.

2.4 Discussion

In this section, I completed three major proofs and five example problems. I carefully chose the three proofs and their related theorems/propositions because they each directly applied

in the example problems and in later sections of the portfolio. I chose example problem 1 because I used the definition of linear independence in a proof, an important technique in solving problems in linear algebra. I included example problems 2 and 3 because they both argued using dimensionality and directly applied the Fundamental Theorem of Linear Maps, another important tool in problem solving. Finally, I chose problem 4 because it gives an introduction to the structure of the matrices of linear, something I explore in sections 4 and 5. Problem 5 was finally chosen because it gives an example of a linear map (the differentiation operator), and demonstrates a change of basis problem, an important technique used in representing the behavior of linear maps.

3 Orthogonal Complements and Minimization Problems

I find minimization problems quite interesting because they have a lot of real-world applications in areas such as machine learning and economics. In this section, I will explore some proofs and nice problems in this area.

3.1 Orthogonal Complements

To begin, I'd like to introduce the notion of an orthogonal complement of a subspace.

Definition 5. Suppose we have a vector space V over a scalar field \mathbb{F} , where V has some inner product $\langle \cdot, \cdot \rangle$ defined (that is, V is an inner product space), and a subspace U of V . Then, the orthogonal complement of U , denoted U^\perp , is defined by

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

In words, every vector in U^\perp is orthogonal to every vector in U .

3.1.1 Direct Sum of a Subspace and Its Orthogonal Complement

A result that may seem intuitive at first is the fact that if we assume that V is finite-dimensional (and hence U and U^\perp are as well),

$$U \oplus U^\perp = V.$$

That is, V is a direct sum of any subspace and its orthogonal complement. The proof has a bit more intricacy, however. The proof has two parts: showing that $V = U + U^\perp$, and showing that $U \cap U^\perp = \{0\}$. Both of these conditions being true will imply that the sum is direct, which follows from the definition of a direct sum.

Proof. Suppose V is a finite-dimensional vector space. Next, pick some vector $v \in V$. Since V is finite dimensional, then we can pick some subspace U of V which is also finite dimensional. So, we can pick some basis for U of length $m = \dim U$. We are dealing with orthogonality, so it is nice to pick an orthonormal basis for U , which we know exists because we can always perform Gram-Schmidt on any linearly independent set of vectors to get an orthonormal list that spans the same space as the original set. Let's pick u_1, \dots, u_m to be this orthonormal basis of U . Next, let some vector $u = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m$. Clearly, $u \in U$, as it is a linear combination of the basis vectors.

We can state that $v = v + (u - u)$, so $v = u + v - u$, as we are just adding zero. Next, let $w = v - u$. Now, we can note that since our basis for U , u_1, \dots, u_m , are an orthonormal list, for $j = 1, \dots, m$,

$$\begin{aligned}\langle w, u_j \rangle &= \langle v - u, u_j \rangle \\ &= \langle v, u_j \rangle - \langle u, u_j \rangle \\ &= \langle v, u_j \rangle - \langle \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m, u_j \rangle \\ &= \langle v, u_j \rangle - (\langle v, u_j \rangle) \langle u_j, u_j \rangle \\ &= \langle v, u_j \rangle - \langle v, u_j \rangle \\ &= 0\end{aligned}$$

If this holds for all basis vectors chosen for U , then this will hold for any such chosen vector $u \in U$, and hence v . So, w is going to be orthogonal to any vector in U , and by definition, $w \in U^\perp$. We defined w as $w = v - u$, so we have that $v = u + w$, where $u \in U$, $w \in U^\perp$, and it must be that $V = U + U^\perp$. Indeed, V is a sum of a subspace and its orthogonal complement. But is it *direct*?

We will now show that $U \cap U^\perp = \{0\}$. If we pick some $v \in U \cap U^\perp$, then it must be the case that $\langle v, v \rangle = 0$ because v is in both the subspace U and its orthogonal complement U^\perp . The only vector which satisfies this property is $v = 0$, so indeed $U \cap U^\perp = \{0\}$.

Since we have that $V = U + U^\perp$ and $U \cap U^\perp = \{0\}$, then $V = U \oplus U^\perp$, completing the proof. \square

Now, you might ask yourself: if we know that for a finite-dimensional V , $V = U + U^\perp$, is it the case that $V - U = U^\perp$? Can we simply perform a set difference to obtain the orthogonal complement? In short, no! The key point here is that U^\perp is also a subspace. Performing the set difference will exclude the zero vector from $V - U$, as we defined U as a subspace, and hence contains the zero vector. So, $V - U \neq U^\perp$. This example briefly highlights the behavior of subspaces, and how they differ from regular subsets.

Example Problem 1: (6.C #2) Suppose U is a finite-dimensional subspace of V . Prove that $U^\perp = \{0\}$ if and only if $U = V$.

Solution: For this problem, I will proceed with a direct proof going in both directions of the biconditional.

Proof. (\rightarrow) If $U^\perp = \{0\}$, and we know that $V = U \oplus U^\perp$, we can pick some vector $v \in V$ by picking $u \in U$ and $0 \in U^\perp$, and writing $v = u + 0 = u$. Since $v = u$, then any vector in V is also a vector in U , and therefore $V \subseteq U$. Since U is a subspace of V , then $U \subseteq V$, so $V = U$.

(\leftarrow) Suppose $V = U$, and we pick any vector $u \in U^\perp \subset V = U$. By definition of the orthogonal complement, u will be orthogonal to every vector in $U = V$. Since $u \in U$ is also true, u will be orthogonal to itself. The only vector which is orthogonal to itself is the zero vector, so $u = 0$ is the only choice of $u \in U^\perp$ we can make. So, $U^\perp = \{0\}$. \square

3.2 Minimization Problems and How They Relate to Orthogonal Complements

Minimization problems in linear algebra are a direct consequence of the existence of orthogonal complements of subspaces, and in this section we will see why. First, we will begin with a definition.

Definition 6. Suppose V is an inner product space and U is a finite-dimensional subspace of V . Then, we define a new operator $P_U \in \mathcal{L}(V)$, named the orthogonal projection operator, which projects any vector $v \in V$ to a vector in U . We define P_U by taking any $v \in V$, writing $v = u + w$, where $u \in U$ and $w \in U^\perp$ (recall the proof in section 3.1.1), and setting $P_U v = u$.

Now, suppose we have some orthonormal basis for U composed of u_1, \dots, u_m . We can extend this basis to form a basis of V by adding vectors w_1, \dots, w_l , which in fact are a basis of U^\perp . Now, we can write any vector $v \in V$ as

$$v = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m + \langle v, w_1 \rangle w_1 + \dots + \langle v, w_l \rangle w_l.$$

Now, let $u \in U$ be

$$u = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m,$$

and $w \in U^\perp$ be

$$w = \langle v, w_1 \rangle w_1 + \dots + \langle v, w_l \rangle w_l.$$

We now have that $v = u + w$ for $u \in U$, $w \in U^\perp$, and then by equating the coefficients of u and v , we see that $u = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m$. So, we can write our orthogonal projection operator for any $v \in V$, given some orthonormal basis of U , as

$$P_U v = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m.$$

This is a very helpful result for problems relating to computation.

3.2.1 Minimization with the Orthogonal Projection

What we are interested in doing is minimizing the distance to a subspace. Suppose we are working with some inner-product vector space V , and one of its subspaces U . We pick some vector $v \in V$, and our goal is to find some vector $u \in U$ which minimizes the distance from v to u . What we mean by *distance* is the norm of the difference between v and u . That is, we are interested in finding

$$\arg \min_u \|v - u\|.$$

A neat result which helps us achieve this is that for any $v \in V$ and any $u \in U$,

$$\|v - P_U v\| \leq \|v - u\|.$$

Curiously, the vector in U which minimizes the distance is the orthogonal projection of v onto U ! That is, $\arg \min_u \|v - u\| = P_U v$. Let's prove this result.

Proof. The squared norm is typically easier to work with, so let's begin with the fact that

$$\|v - P_U v\|^2 \leq \|v - P_U v\|^2 + \|P_U v - u\|^2.$$

This holds because $\|P_U v - u\|^2 \geq 0$, as norms are always non-negative. Next, let's recall that by the definition of the orthogonal projection operator, $P_U v = u'$ where $v = u' + w$, for $u' \in U$ and $w \in U^\perp$. Then, $v - P_U v = (u' + w) - u' = w \in U^\perp$. Additionally, $P_U v - u = u' - u \in U$, as both u' and u are in U . So, it is the case that $\langle v - P_U v, P_U v - u \rangle = 0$ (they are orthogonal to each other), so by the Pythagorean theorem,

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

We see then that

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

so by transitivity,

$$\|v - P_U v\|^2 \leq \|v - u\|^2.$$

Taking the square root of both sides we get

$$\|v - P_U v\| \leq \|v - u\|,$$

which completes our proof. \square

Let's now take a look at some example problems. This first problem is proof based, and relies on the proof from section 3.1.1.

Example Problem 2: (6.C #5) Suppose V is a finite-dimensional inner product space and U is a subspace of V . Show that $P_{U^\perp} = I - P_U$, where I is the identity operator on V .

Solution: I will proceed with a direct proof.

Proof. Let u_1, \dots, u_m be an orthonormal basis for U . Next, suppose that w_1, \dots, w_l form an orthonormal basis for U^\perp . Next, pick any $v \in V$. Then,

$$P_{U^\perp} v = \langle v, w_1 \rangle w_1 + \dots + \langle v, w_l \rangle w_l.$$

We can add $0 = v - v$ to both sides to get

$$P_{U^\perp} v = v - v + \langle v, w_1 \rangle w_1 + \dots + \langle v, w_l \rangle w_l.$$

Since we gave orthonormal bases for U and U^\perp , and since $V = U \oplus U^\perp$ as V is finite-dimensional, then we have a basis for V : $u_1, \dots, u_m, w_1, \dots, w_l$. We can thus write $v \in V$ using this orthonormal basis as

$$v = \langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m + \langle v, w_1 \rangle w_1 + \dots + \langle v, w_l \rangle w_l.$$

So, we have that

$$P_{U^\perp} v = v - (\langle v, u_1 \rangle u_1 + \dots + \langle v, u_m \rangle u_m),$$

which is

$$P_{U^\perp} v = v - P_U v.$$

So, we see that $P_{U^\perp} v = (I - P_U)v$, and so $P_{U^\perp} = I - P_U$. \square

This next question is more computationally heavy.

Example Problem 3: (6.C #12) Find $p \in \mathcal{P}_3(\mathbb{R})$ such that $p(0) = 0$, $p'(0) = 0$, and

$$\int_0^1 |2 + 3x - p(x)|^2 dx$$

is as small as possible.

Solution: Let's first define the inner product between $f, g \in \mathcal{P}_3(\mathbb{R})$ as

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx.$$

Refer to appendix A.2 for a proof of why this is an inner product. With this inner product defined we can observe that

$$\int_0^1 |2 + 3x - p(x)|^2 dx = \|(2 + 3x) - p(x)\|^2$$

under the definition of our inner product. So, our problem is reformulated to finding a p such that $p(0) = 0$, $p'(0) = 0$, and $\|(2 + 3x) - p(x)\|$ is as small as possible. Thus, we have a minimization problem of the same structure as before. Let's consider the subspace U of $\mathcal{P}_3(\mathbb{R})$ which is $U = \{p \in \mathcal{P}_3(\mathbb{R}) : p(0) = 0, p'(0) = 0\}$. Our minimization problem now requires us to find $p(x) = P_U(2 + 3x)$, as found in the proof from section 3.2.1.

Let's compute an orthonormal basis for U . First, a basis for U is x^2, x^3 . The constant term $p(x) = a$ is not included because it does not satisfy $p(0) = 0$. The degree 1 term $p(x) = ax$, for $a \neq 0, a \in \mathbb{R}$, is not included because $p'(0) = a \neq 0$. Now we perform Gram-Schmidt to get an orthonormal basis. Normalize x^2 by defining $e_1 = \sqrt{5}x^2$. This is normalized because

$$\langle e_1, e_1 \rangle = \int_0^1 (\sqrt{5}x^2)(\sqrt{5}x^2)dx = 5 \int_0^1 x^4 dx = 1.$$

Now, define

$$e_2 = \frac{x^3 - \langle x^3, e_1 \rangle e_1}{\|x^3 - \langle x^3, e_1 \rangle e_1\|}.$$

The numerator is

$$x^3 - \langle x^3, e_1 \rangle e_1 = x^3 - \left(\int_0^1 (x^3)(\sqrt{5}x^2)dx \right) \sqrt{5}x^2 = x^3 - (5/6)x^2.$$

The denominator is

$$\|x^3 - (5/6)x^2\| = \sqrt{\int_0^1 (x^3 - (5/6)x^2)^2 dx} = 1/(6\sqrt{7}).$$

So, $e_2 = 6\sqrt{7}(x^3 - (5/6)x^2) = \sqrt{7}(6x^3 - 5x^2)$.

To wrap up, we now have an orthonormal basis for U : $e_1 = \sqrt{5}x^2$ and $e_2 = \sqrt{7}(6x^3 - 5x^2)$. Now we finally can compute $p(x) = P_U(2 + 3x)$, which is our desired minimizing vector:

$$\begin{aligned}
P_U(2 + 3x) &= \langle 2 + 3x, e_1 \rangle e_1 + \langle 2 + 3x, e_2 \rangle e_2 \\
&= \left(\int_0^1 (2 + 3x) \sqrt{5} x^2 dx \right) \sqrt{5} x^2 \\
&\quad + \left(\int_0^1 (2 + 3x) (\sqrt{7}(6x^3 - 5x^2)) dx \right) (\sqrt{7}(6x^3 - 5x^2)) \\
&= 5 \left(\int_0^1 (2 + 3x) x^2 dx \right) x^2 \\
&\quad + 7 \left(\int_0^1 (2 + 3x) ((6x^3 - 5x^2)) dx \right) (6x^3 - 5x^2) \\
&= 5(17/12)x^2 + 7(-29/60)(6x^3 - 5x^2) \\
&= (85/12)x^2 - (203/60)(6x^3 - 5x^2) \\
&= 24x^2 - (203/10)x^3
\end{aligned}$$

To conclude this problem, the $p \in U$ which minimizes $\int_0^1 |2 + 3x - p(x)|^2 dx$ is $p(x) = 24x^2 - (203/10)x^3$

3.3 Discussion

In this section I gave two proofs and three example problems relating to orthogonal complements and minimization problems. I chose the proof in section 3.1.1 because it is used in many applications of orthogonal complements, and it was directly applied in example problems 1 and 2. I chose the second proof because it directly leads to the most important result in minimization problems, which is that the vector that minimizes the norm in minimization problems is the orthogonal projection onto the desired subspace. The subspace in these minimization problems is effectively the constraint on our solution, which is a pivotal part of minimization problems in general. Example problem 2 combined ideas from both proofs, while I included example problem 3 because it served as a practical application of the definitions and proofs in the entirety of this section.

4 Eigenvectors and Eigenvalues

Eigenstuff are fundamental concepts in all of linear algebra. In a way, they provide structure to linear operators. Additionally, they are a building block to determining if an operator has an upper triangular or diagonal matrix under some choice of basis, which come in handy for applications. This section will explore a collection of definitions, proofs, and problems related to this overarching topic.

4.1 Basics of Eigenvalues and Eigenvectors

Definition 7. Suppose V is a vector space over a scalar field \mathbb{F} and $T \in \mathcal{L}(V)$ – that is, T is a linear operator over V . A number $\lambda \in \mathbb{F}$ is an eigenvalue of T if there exists a non-zero vector $v \in V$ such that $Tv = \lambda v$.

An important detail is that $v \neq 0$, because otherwise any $\lambda \in \mathbb{F}$ satisfies $T0 = \lambda 0$. Geometrically we can think of λ being some scaling factor for some $v \in V$. However, not all $v \in V$ satisfy this definition. The ones that do are called eigenvectors.

Definition 8. Suppose V is a vector space over a scalar field \mathbb{F} and $T \in \mathcal{L}(V)$. We call a vector $v \in V$ an eigenvector associated to some eigenvalue $\lambda \in \mathbb{F}$ of T if $v \neq 0$ and $Tv = \lambda v$.

Example Problem 1: (5.A #7) Suppose $T \in \mathcal{L}(\mathbb{R}^2)$ is defined by $T(x, y) = (-3y, x)$. Find the eigenvalues of T , if any.

Solution: Suppose that λ is an eigenvalue of T . Then, there exists some $v \in \mathbb{R}^2$, such that $v \neq 0$ and $Tv = \lambda v$. Setting $v = (x, y)$, we need to find λ such that $T(x, y) = \lambda(x, y) = (\lambda x, \lambda y)$, and $x \neq 0, y \neq 0$. Since $T(x, y) = (-3y, x)$, we need $\lambda x = -3y$ and $\lambda y = x$. This implies that $\lambda^2 y = -3y$, and since $y \neq 0$, $\lambda^2 = -3$. Clearly, our operator has no real eigenvalues. We are working with $T \in \mathcal{L}(\mathbb{R}^2)$, so the scalar field is \mathbb{R} . Definition 7 requires that the eigenvalues belong to the same scalar field, and thus, our example of T does not have any eigenvalues.

Now, suppose we extend our operator to instead work in a complex field. That is, suppose $T \in \mathcal{L}(\mathbb{C}^2)$. Then, we allow for $\lambda_1 = \sqrt{3}i$ and $\lambda_2 = -\sqrt{3}i$ to be the eigenvalues of T .

An important theorem is the fact that eigenvectors that correspond to distinct eigenvalues are linearly independent. Suppose we have some vector space V , defined as above, and $T \in \mathcal{L}(V)$. Then, let $\lambda_1, \dots, \lambda_m$ be the m distinct eigenvalues of T , and v_1, \dots, v_m be corresponding eigenvectors. Then, it is the case that v_1, \dots, v_m are linearly independent. This property will be proven with a contradiction.

Proof. Suppose that the converse of the statement is true. That is, suppose that the eigenvectors v_1, \dots, v_m are linearly dependent. Then, we can pick the smallest positive integer k such that

$$v_k \in \text{span}(v_1, \dots, v_{k-1}),$$

which exists by the linear dependence lemma. Here, v_1, \dots, v_{k-1} is the smallest linearly independent list of vectors we can choose. By the definition of linear dependence, there exists scalars $a_1, \dots, a_{k-1} \in \mathbb{F}$ such that

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

Applying T to both sides gives

$$Tv_k = T(a_1 v_1 + \dots + a_{k-1} v_{k-1}).$$

Now, since each v_1, \dots, v_k are eigenvectors, they are non-zero vectors, this evaluates to

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

Additionally, we can state that

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

where we multiplied both sides of our equation for v_k by λ_k . Thus, we have that

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

and simplifying this gives

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

We stated that v_1, \dots, v_{k-1} are linearly independent, so each coefficient of each v_1, \dots, v_{k-1} is zero. Since we stated that $\lambda_1, \dots, \lambda_k$ are distinct, it must be the case that $a_1 = \cdots = a_{k-1} = 0$. This leads us to a contradiction, because this means that $v_k = 0v_1 + \cdots + 0v_{k-1} = 0$, which is not true for an eigenvector. So, it must be the case that our initial assumption is false, and thus v_1, \dots, v_m are linearly independent. \square

This proof leads us to the following problem:

Example Problem 2: (5.A #29) Suppose $T \in \mathcal{L}(V)$, where V is a vector space, and $\dim \text{range } T = k$. Prove that T has at most $k + 1$ distinct eigenvalues.

Solution: For this problem, we proceed with a direct proof.

Proof. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T . We are looking to prove that $m \leq k + 1$. Now, let v_1, \dots, v_m be eigenvectors associated with each eigenvalue. From the proof in section 4.1, these vectors must be linearly independent. Also, for an eigenvector v_j to be in the range of T (for $j = 1, \dots, m$), it must be that its associated eigenvalue, λ_j , be non-zero. This is because $v_j \in \text{range } T$ implies that there is some vector $u \in V$ such that $Tu = v_j$. First, we note that $Tv_j = \lambda_j v_j$, so if we assume $\lambda_j \neq 0$, $T(v_j/\lambda_j) = v_j$. Hence, we have $u = v_j/\lambda_j$. Clearly, $v_j \in \text{range } T$ if $\lambda_j \neq 0$.

Now, since $\dim \text{range } T = k$, we can span $\text{range } T$ with k linearly independent of the eigenvectors. So, we could pick at most k out of the m eigenvectors to form a basis for $\text{range } T$. This, along with the previous fact, implies that we can have at most k distinct non-zero eigenvalues for T . We can add on zero as an eigenvalue, with the only catch that its associated eigenvector is not in the range of T , resulting in at most $m = k + 1$ distinct eigenvalues, and thus for any operator T with $\dim \text{range } T = k$ and m distinct eigenvalues, $m \leq k + 1$. This completes our proof. \square

4.2 Upper Triangular and Diagonal Matrices

Upper triangular and diagonal matrices are useful in the applications of linear operators. If a linear operator has an upper triangular or diagonal matrix with respect to some choice of basis, then computations with them become easier.

4.2.1 Upper Triangular Matrices of Operators

Definition 9. A matrix is upper triangular if all entries below the diagonal are zero.

Suppose V is a vector space over a scalar field \mathbb{F} and v_1, \dots, v_n is a basis for V . Let $T \in \mathcal{L}(V)$. With this given, then the following three points are equivalent:

- (a) A matrix of T with respect to v_1, \dots, v_n is upper triangular.
- (b) $Tv_j \in \text{span}(v_1, \dots, v_j)$, for each $j = 1, \dots, n$.
- (c) $\text{span}(v_1, \dots, v_j)$ is invariant under T for each $j = 1, \dots, n$.

Another important result applies when $\mathbb{F} = \mathbb{C}$. In this case, it is a fact that every operator on a finite-dimensional complex vector space has an upper triangular matrix with respect to some choice of basis for V . Refer to propositions 5.26 and 5.27 in *Linear Algebra Done Right* for full proofs of these facts.

Example Problem 3: (5.B #20) Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Prove that T has an invariant subspace of dimension k for each $k = 1, \dots, \dim V$.

Solution: I will proceed with a direct proof.

Proof. If V is a finite-dimensional complex vector space, then $T \in \mathcal{L}(V)$ has an upper triangular matrix with some choice of basis. Let this choice of basis be v_1, \dots, v_n . Hence, $n = \dim V$. Now, we will define n k -dimensional subspaces of V , whose basis are formed by picking the first k basis vectors of V . That is, we will choose the following n subspaces, denoted U_k for $k = 1, \dots, n$.

$$U_1 = \text{span}(v_1)$$

$$U_2 = \text{span}(v_1, v_2)$$

$$U_3 = \text{span}(v_1, v_2, v_3)$$

...

$$U_n = \text{span}(v_1, \dots, v_n) = V$$

Also, since V has an upper triangular matrix with respect to this choice of basis, we know that $Tv_1 = a_{1,1}v_1$, $Tv_2 = a_{1,2}v_1 + a_{2,2}v_2$, and so on, until $Tv_n = a_{1,n}v_1 + \dots + a_{n,n}v_n$, for each $a_{i,j} \in \mathbb{F}$ for $i = 1, \dots, n$, $j = 1, \dots, n$. We can then notice that $Tv_1 \in U_1$, $Tv_2 \in U_2$, and so on, until $Tv_n \in U_n$. Thus, each U_1, \dots, U_n are our desired invariant subspaces, and there is one of each dimensionality for $k = 1, \dots, n$.

We can briefly show that they are invariant by picking some vector $u_k \in U_k$ where $u_k = c_1v_1 + \dots + c_kv_k$, where $1 \leq k \leq n$ and $c_1, \dots, c_k \in \mathbb{C}$. Then,

$$Tu_k = T(c_1v_1 + \dots + c_kv_k) = c_1Tv_1 + \dots + c_kTv_k.$$

We defined each U_j for $j = 1, \dots, k$ such that $U_1 \subset U_2 \subset \dots \subset U_k$, and each $Tv_1 \in U_1, \dots, Tv_k \in U_k$, so $Tv_1 \in U_k, \dots, Tv_k \in U_k$. Since vector spaces are closed under addition and scalar multiplication, then $Tu_k = c_1Tv_1 + \dots + c_kTv_k \in U_k$, as desired. \square

4.2.2 Diagonal Matrices of Operators

Definition 10. A matrix is diagonal if all entries above and below the diagonal are zero.

A diagonal matrix is also upper triangular. So, we require stronger equivalent conditions to those in section 4.2.1: Suppose V is a vector space, and let $T \in \mathcal{L}(V)$, such that $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T . Then, the following conditionals are equivalent:

- (a) T is diagonalizable (T has a diagonal matrix with respect to some choice of basis of V).
- (b) V has a basis of eigenvectors of T .
- (c) There exist n 1-dimensional subspaces of V , where $n = \dim V$, such that $V = U_1 \oplus \dots \oplus U_n$, such that each U_1, \dots, U_n are invariant under T .
- (d) $V = E(\lambda_1, T) + \dots + E(\lambda_m, T)$, where $E(\lambda_k, T) = \text{null}(T - \lambda_k I)$ is the eigenspace associated with an eigenvalue λ_k , for $k = 1, \dots, m$.
- (e) $\dim V = \dim E(\lambda_1, T) + \dots + \dim E(\lambda_m, T)$.

Although I will not prove the equivalent conditions, I will apply them in a problem, and they are definitely worth mentioning (for a full proof, see theorem 5.41 in *Linear Algebra Done Right*).

Additionally, a consequence of these conditions is that if $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable. This is a useful result because it allows us to quickly identify if an operator is diagonalizable. I will show this result using a direct proof.

Proof. If T has $\dim V$ distinct eigenvalues, then by the proof in section 4.1, we have that T has $\dim V$ linearly independent eigenvectors. We now have a linearly independent list of $\dim V$, which directly implies that the eigenvectors of T form a basis V , by the definition of a basis. By point (b), this implies that T is diagonalizable. \square

This property can also be thought of intuitively, without the equivalent conditions, and simply understanding how a matrix is defined. Given our understanding of eigenvectors and eigenvalues, we know that if we have $n = \dim V$ linearly independent eigenvectors associated with n distinct eigenvalues, V must have a basis of these eigenvectors. Let v_1, \dots, v_n be this basis. Then, $Tv_1 = \lambda_1 v_1, \dots, Tv_n = \lambda_n v_n$, where $\lambda_1, \dots, \lambda_n$ are the distinct eigenvalues. If we define a matrix under this basis, then clearly we will only see entries along the diagonal, and these entries are $\lambda_1, \dots, \lambda_n$. The entries of a diagonal matrix are nothing but the eigenvalues of the corresponding operator.

Example Problem 4: (5.C #6) Suppose V is a finite-dimensional vector space, $T \in \mathcal{L}(V)$, and T has $\dim V$ distinct eigenvalues, and $S \in \mathcal{L}(V)$ has the same eigenvectors (but not necessarily the same eigenvalues). Prove that $ST = TS$.

Solution: We proceed with a direct proof.

Proof. If T has $\dim V$ distinct eigenvalues, then T is diagonalizable with respect to some choice of basis. We can see that this choice of basis is in fact a basis of eigenvectors, from the equivalent conditions above. Let this basis of eigenvectors be v_1, \dots, v_n , where $n = \dim V$, and the distinct eigenvalues be $\lambda_1, \dots, \lambda_n$. Now, since S shares eigenvectors, S must also have n distinct eigenvalues, denoted μ_1, \dots, μ_n . Suppose we pick some vector $u \in V$ written in terms of the basis of eigenvectors. We can write $u = c_1 v_1 + \dots + c_n v_n$, where $c_1, \dots, c_n \in \mathbb{F}$. Then,

$$\begin{aligned} STu &= S(Tu) \\ &= S(c_1 \lambda_1 v_1 + \dots + c_n \lambda_n v_n) \\ &= c_1 \mu_1 \lambda_1 v_1 + \dots + c_n \mu_n \lambda_n v_n \\ &= c_1 \lambda_1 (\mu_1 v_1) + \dots + c_n \lambda_n (\mu_n v_n) \\ &= c_1 \lambda_1 S v_1 + \dots + c_n \lambda_n S v_n \\ &= c_1 T(S v_1) + \dots + c_n T(S v_n) \\ &= TS(c_1 v_1 + \dots + c_n v_n) \\ &= TSu \end{aligned}$$

Thus, $STu = TSu$ for all $u \in V$, so $ST = TS$. □

4.3 Discussion

In this section, I completed two proofs and four example problems. I chose the problems and proof because they best represented the notion that *eigenstuff* can be used to describe the structure of linear operators. I chose the first problem as it served as both an example and non-example of eigenvalues, and showing the flexibility that complex operators offer. I completed the first proof because it is an extremely important result that demonstrates the structure of linear operators very well. The results of this proof come back when I discussed upper triangular and diagonal matrices of operators. I chose the second and third problems because they used dimensionality arguments in proofs, a very important skill in linear algebra, and because they directly applied the theorems and proofs. The third and fourth problem directly show how eigenvectors and eigenvalues provide structure to operators through invariant subspaces.

5 Markov Chains

A difference equation is a recurrence relation defined by

$$x_{t+1} = T x_t,$$

where $T \in \mathcal{L}(V)$, and $x_t, x_{t+1} \in V$, for all integers $t \geq 0$. In this section of my portfolio I will explore a particular set of difference equations: Markov Chains. A Markov Chain is a difference equation that deals with probabilities, so we work with real vector spaces, i.e, $V = \mathbb{R}^n$ for positive integers n .

Let's define some core terminology required to study Markov Chains:

Definition 11. A vector $x_t \in \mathbb{R}^n$ is called a *state vector* if and only if:

- (a) Each entry in x_t is non-negative.
- (b) The sum of all entries in x_t is 1.

A state vector represents the state of a system at some integer time t [2]. Under this definition, a state vector is also called a probability vector.

Each entry in a state vector can be interpreted as the probability of being in a particular component of a state. For example, suppose we define a system as a rat in a maze with three rooms, and with no exit or entry. We can encode the state of the system at time t through $x_t \in \mathbb{R}^3$, where the first entry of x_t is the probability of being in the first room, the second entry being the probability of being in the second room, and the third entry being the probability of being in the third room. The rat will always be in the maze, so the sum of all entries must be 1. A probability cannot be non-negative, which motivates points (a) and (b) above.

I will denote the set of all state vectors with n entries as \mathbb{S}^n . It is important to note that \mathbb{S}^n is not a subspace of \mathbb{R}^n , as it does not contain the zero vector.

Definition 12. An operator $T \in \mathcal{L}(\mathbb{R}^n)$ is called a *transition operator* if it maps a state vector to another state vector. That is, for all $v \in \mathbb{R}^n$, if $v \in \mathbb{S}^n$, then $Tv \in \mathbb{S}^n$.

A transition operator describes how the state of a system changes from one time step to the next. Section 5.1 will explore a practical example.

5.1 A Practical Introduction to Markov Chains

Markov Chains have practical applications in areas such as economics, biology, and computer science. Let's look at a concrete example to motivate their study. Suppose we are interested in seeing how the movement of people between a city and its suburbs evolves each year, under the assumption that the total population is constant (that is, the population only stays in either the city or suburbs, there are no births or deaths, etc.). Let $x_t \in \mathbb{R}^2$ be the state of the system at year $t = 0, 1, 2, \dots$, where $x_t = (p_{\text{city},t}, p_{\text{suburbs},t})$ and $p_{\text{city},t}, p_{\text{suburbs},t}$ denote the proportion of the total population in the city and suburbs at year t . Naturally, $p_{\text{city},t} + p_{\text{suburbs},t} = 1$, and $p_{\text{city},t}, p_{\text{suburbs},t} \geq 0$, because proportions cannot be negative. Our vector x_t is a state vector. We can define our transition operator $T \in \mathcal{L}(\mathbb{R}^2)$; T describes how some vector in \mathbb{R}^2 changes from one year to the next. Now, let's define

$$x_{t+1} = T(p_{\text{city},t}, p_{\text{suburbs},t}) = (0.95p_{\text{city},t} + 0.1p_{\text{suburbs},t}, 0.05p_{\text{city},t} + 0.9p_{\text{suburbs},t}).$$

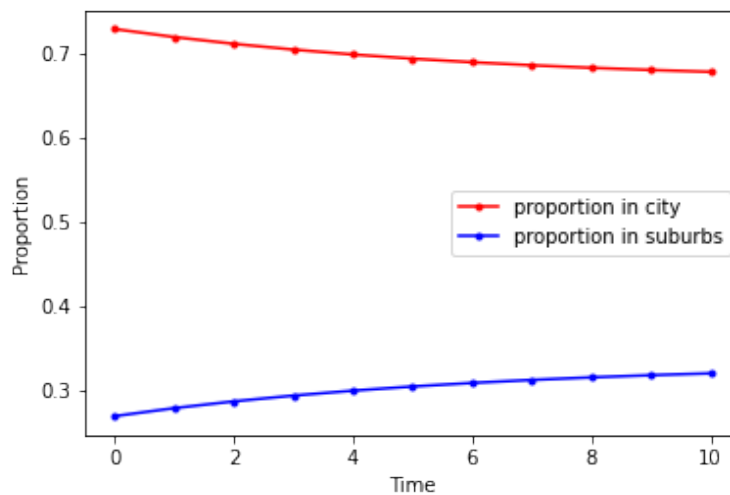
We can describe what this means in English by looking at the first and second entries of the resulting vector: The first slot tells us the proportion of people in year $t + 1$ that will live in the city, and the second slot tells us the proportion of people in year $t + 1$ that will live in the suburbs. Looking at the first slot, we will have 95% of people in the city stay in the city, and 10% of people from the suburbs move to the city. Looking at the second slot, we will have the remaining 5% of city people move to the suburbs, and the remaining 90% of suburb-dwellers staying in the suburbs.

T is a transition operator because its entries sum to one. We have that

$$\begin{aligned} p_{\text{city},t+1} + p_{\text{suburbs},t+1} &= 0.95p_{\text{city},t} + 0.1p_{\text{suburbs},t} + 0.05p_{\text{city},t} + 0.9p_{\text{suburbs},t} \\ &= p_{\text{city},t} + p_{\text{suburbs},t} \\ &= 1, \end{aligned}$$

as desired. The entries are also non-negative because both $p_{\text{city},t}, p_{\text{suburbs},t} \geq 0$ and the coefficients are non-negative.

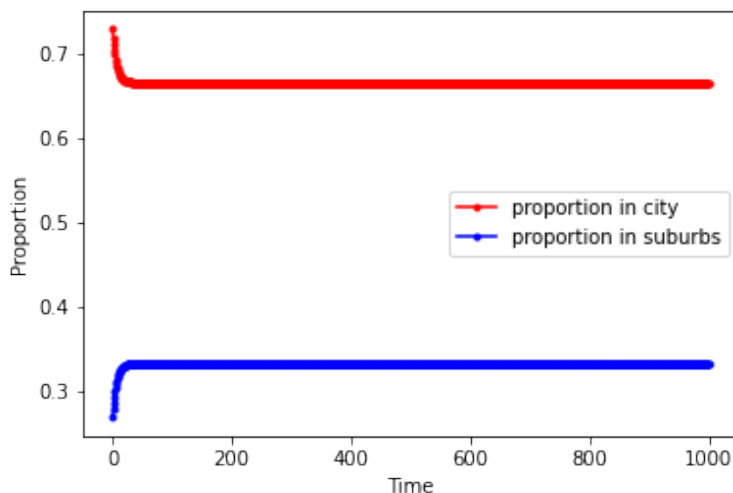
Suppose that at year $t = 0$, 73% of the population lives in the city, and the remaining 27% live in the suburbs. What will the population be in year $t = 1$? This found by computing $x_1 = Tx_0$, where $x_0 = (0.73, 0.27)$. So, $x_1 = T(0.73, 0.27) = (0.95 \cdot 0.73 + 0.1 \cdot 0.27, 0.05 \cdot 0.73 + 0.9 \cdot 0.27) = (0.7205, 0.2795)$. We found that $x_1 = (0.7205, 0.2795)$. Now, what if we wanted to compute x_{10} ? We can just repeatedly compute $x_2 = Tx_1, x_3 = Tx_2, \dots, x_{10} = Tx_9$, which is equivalent to computing $x_{10} = T^{10}x_0$ – we repeatedly apply the same transition operator on the initial state! Lets visualize this repeated operation for each time step $t = 0, \dots, 10$:



5.2 Steady States

5.2.1 Steady State Vectors

The behavior of a Markov Chain as it tends to infinity is entirely dependent on the transition operator. Let's consider the same example as in section 5.1, but increasing our maximum time step in our plot to some absurdly high number in the context of the problem, say 1000. How does our system change over a thousand years?



There seems to be some limiting behavior in our example, and it may appear that we are reaching some state $q \in \mathbb{R}^2$ such that

$$\lim_{n \rightarrow \infty} T^n x_0 = q.$$

For certain transition operators, such a state exists. We call such a vector q a *steady state*. Intuitively, we can see that $Tq = q$ for a steady state q , so we note that q is an eigenvector of T associated with the eigenvalue $\lambda = 1$. For a steady state, repeated application of the transition operator does not change the state. For this particular example, the eigenvectors associated with $\lambda = 1$ for T are $\text{span}((2, 1))$. However, we need the steady state to be a state vector. That is, the entries sum to 1 and the entries are non-negative. We simply need to find the factor which normalizes the sum and keeps entries positive, which is $1/(2+1) = 1/3$. So, we have our steady-state vector $q = (2/3, 1/3)$. The steady state can be seen in the plot above, where the red line approaches $2/3$ and the blue line approaches $1/3$.

5.2.2 The Matrix of a Transition Operator

Suppose we have a transition operator $T \in \mathcal{L}(\mathbb{R}^n)$. An important question to ask is whether a transition operator has a steady state vector, or if the steady state is unique. To begin answering this question, we must first introduce stochastic matrices, also known as Markov matrices.

Definition 13. Let A be a square matrix with real entries. A is a stochastic matrix if each of its entries are non-negative and the sum of the entries in each row is equal to 1.

Let's explore what the matrices of any transition operator $T \in \mathbb{R}^n$ look like. Suppose $x = (x_1, \dots, x_n)$ is a state vector in \mathbb{S}^n . We can write Tx in the standard form

$$T(x_1, \dots, x_n) = (a_{1,1}x_1 + \dots + a_{1,n}x_n, \dots, a_{n,1}x_1 + \dots + a_{n,n}x_n)$$

for coefficients $a_{i,j} \in \mathbb{R}, 1 \leq i, j \leq n$. We need a basis of \mathbb{R}^n to construct a matrix of T , so let's pick the standard basis e_1, \dots, e_n . The standard basis of \mathbb{R}^n is trivially a set of state

vectors under our definition, so let's use them to construct our matrix of T , $\mathcal{M}(T)$:

$$\mathcal{M}(T) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix}$$

Now, I claim that $\mathcal{M}(T)$ is a stochastic matrix. Here is a proof:

Proof. First, we prove that each column of $\mathcal{M}(T)$ sum to one. Since T is a transition operator, if $x \in \mathbb{S}^n$, then $Tx \in \mathbb{S}^n$. So, we have that $x_1 + \cdots + x_n = 1$ and for $i = 1, \dots, n$, $x_i \geq 0$. Since $Tx \in \mathbb{S}^n$,

$$a_{1,1}x_1 + \cdots + a_{1,n}x_n + \cdots + a_{n,1}x_1 + \cdots + a_{n,n}x_n = 1$$

and

$$a_{1,1}x_1 + \cdots + a_{1,n}x_n, \dots, a_{n,1}x_1 + \cdots + a_{n,n}x_n \geq 0.$$

Note that because they both sum to one,

$$x_1 + \cdots + x_n = a_{1,1}x_1 + \cdots + a_{1,n}x_n + \cdots + a_{n,1}x_1 + \cdots + a_{n,n}x_n.$$

The right-hand side can be simplified by grouping the coefficients of each x_i , for $i = 1, \dots, n$:

$$x_1 + \cdots + x_n = (a_{1,1} + \cdots + a_{n,1})x_1 + \cdots + (a_{1,n} + \cdots + a_{n,n})x_n.$$

By equating coefficients,

$$a_{1,i} + \cdots + a_{n,i} = 1$$

for all $i = 1, \dots, n$. This proves that all columns of $\mathcal{M}(T)$ sum to one.

Next, we prove that each entry of $\mathcal{M}(T)$ is non-negative. To prove this, assume the contrary. Let's assume that there exists at least one negative entry in $\mathcal{M}(T)$. Suppose $a_{i,j}$ is such a negative entry, where $1 \leq i, j \leq n$. Our matrix then tells us that applying T to the j th basis vector e_j gives

$$Te_j = a_{1,j}e_1 + \cdots + a_{i,j}e_i + \cdots + a_{n,j}e_n$$

which in the standard form of the operator is

$$Te_j = (a_{1,j}, \dots, a_{i,j}, \dots, a_{n,j}).$$

We know that $e_j \in \mathbb{S}^n$, but clearly $Te_j \notin \mathbb{S}^n$, as the resulting vector has a negative entry given by $a_{i,j}$. This is a contradiction, because we defined T as a transition operator. Therefore, each entry in $\mathcal{M}(T)$ must be non-negative.

Therefore, the matrix of any transition operator T under the standard basis is a stochastic matrix. \square

Transition operators have an interesting property regarding their eigenvalues.

Theorem 2. Suppose T is a transition operator. Then:

- (a) 1 is an eigenvalue of T .
- (b) If $\lambda \in \mathbb{R}$ is an eigenvalue of T , then $|\lambda| \leq 1$ [2].

Proof. Suppose $T \in \mathcal{L}(\mathbb{R}^n)$ is a transition operator. Let $M = \mathcal{M}(T)$ under the standard basis of \mathbb{R}^n . We know that M is a stochastic matrix, so its columns sum to one. Now, consider the conjugate transpose, M^* , of M , which is the matrix that corresponds to the adjoint of T , T^* . Since T operates over a real vector space, M will have only real entries, and so will M^* . The rows of M^* will thus sum to one.

Next, consider what happens when we compute the matrix multiplication of M^* with $v = (1 \ \cdots \ 1)^T$ (a column vector with n ones as entries). Suppose

$$M^* = \begin{pmatrix} m_{1,1} & \cdots & m_{1,n} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,n} \end{pmatrix},$$

for $m_{i,j} \in \mathbb{R}$ for all $1 \leq i, j \leq n$. Then,

$$M^*v = \begin{pmatrix} m_{1,1} & \cdots & m_{1,n} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,n} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} m_{1,1} + \cdots + m_{1,n} \\ \vdots \\ m_{n,1} + \cdots + m_{n,n} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

This matrix multiplication represents the linear transformation

$$T^*(e_1 + \cdots + e_n) = e_1 + \cdots + e_n,$$

where e_1, \dots, e_n are the standard basis vectors of \mathbb{R}^n . So, 1 is an eigenvalue of T^* . Now, if 1 is an eigenvalue of T^* , then 1 is also an eigenvalue of T (refer to appendix A.3 for a proof). This proves point (a).

Now, let $\lambda \in \mathbb{R}$ to be any eigenvalue of T . Then, λ is also an eigenvalue of T^* (by the same proof in appendix A.3). Let $x \in \mathbb{R}^n$ be an eigenvector of T^* associated with λ . Then, $T^*x = \lambda x$. We can write x in terms of the standard basis by $x = a_1e_1 + \dots + a_ne_n$ for $a_1, \dots, a_n \in \mathbb{R}$. Then,

$$\begin{aligned} T^*x &= \lambda x = T^*(a_1e_1 + \cdots + a_ne_n) \\ &= a_1T^*e_1 + \cdots + a_nT^*e_n \\ &= a_1(m_{1,1}e_1 + \cdots + m_{n,1}e_n) + \cdots + a_n(m_{1,n}e_1 + \cdots + m_{n,n}e_n) \\ &= (a_1m_{1,1} + \cdots + a_nm_{1,n})e_1 + \cdots + (a_1m_{n,1} + \cdots + a_nm_{n,n})e_n \end{aligned}$$

So, we see that

$$\lambda a_1e_1 + \cdots + \lambda a_ne_n = (a_1m_{1,1} + \cdots + a_nm_{1,n})e_1 + \cdots + (a_1m_{n,1} + \cdots + a_nm_{n,n})e_n.$$

Now, choose the largest a_j such that $|a_j| \geq |a_i|$ for some $j = 1, \dots, n$ and for all $i = 1, \dots, n$ and $i \neq j$. Considering only the j th term in the sum, we see that

$$|\lambda a_j| = |\lambda| |a_j| = |a_1m_{j,1} + \cdots + a_nm_{j,n}| = \left| \sum_{i=1}^n a_im_{j,i} \right|$$

by equating the coefficients of e_j . Then, we have that

$$\left| \sum_{i=1}^n a_im_{j,i} \right| \leq \left| \sum_{i=1}^n m_{j,i} \right| |a_j|,$$

because we chose the maximal a_j . Now, we see that $\sum_{i=1}^n m_{j,i} = 1$, because this is summation over a row of M^* . We thus see that

$$\left| \sum_{i=1}^n m_{j,i} \right| |a_j| = |a_j|,$$

and by transitivity, $|\lambda||a_j| \leq |a_j|$. It is important to note that $|a_j| \neq 0$, because we stated that x was an eigenvector, and hence is non-zero. If the maximal a_j was equal to zero, this would imply that $x = 0$, which is not true. Thus, we can divide both sides of the inequality by $|a_j|$, so $|\lambda| \leq 1$, as desired. This completes the proof of point (b). \square

5.2.3 The Perron-Frobenius Theorem and the Steady State

The Perron-Frobenius theorem finally helps us answer the question of the existence and uniqueness of steady-state vectors of a transition operator $T \in \mathcal{L}(R^n)$.

Theorem 3. The Perron-Frobenius theorem asserts that if A is a square matrix with strictly positive entries, then A has a unique maximal eigenvalue, and it has a corresponding eigenvector with all positive entries. The eigenspace associated with this maximal eigenvalue is one-dimensional [3].

The proof of the theorem is beyond the scope of this portfolio and the course, but it is important to mention.

Combining the results from this theorem and theorem 2, we can see that any transition operator T with a stochastic matrix with strictly positive entries will have a unique maximal eigenvalue $\lambda_{\max} = 1$ with a unique eigenvector with all positive entries. This eigenvector can be scaled to become a state vector, using the same trick from the practical introduction, where we divided the vector by the sum of its entries.

For example, consider the transition operator T from section 5.1. Under the standard basis, its stochastic matrix is

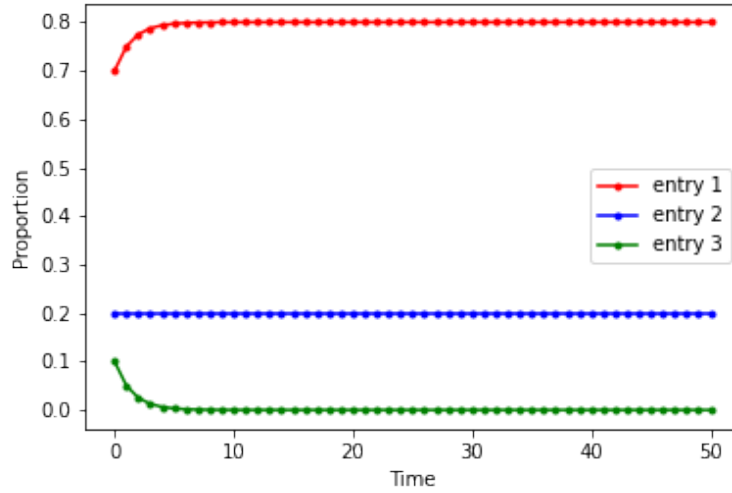
$$\mathcal{M}(T) = \begin{pmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{pmatrix}$$

$\mathcal{M}(T)$ is clearly stochastic, and its entries are strictly positive. So, $(2/3, 1/3)$ is the unique steady state.

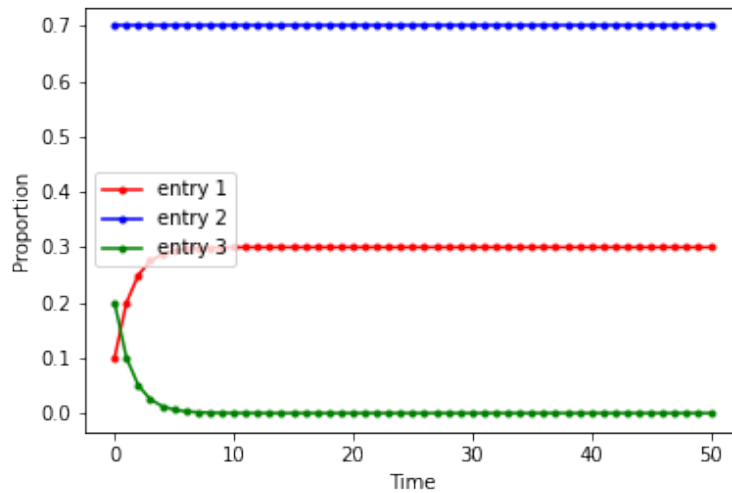
For a counter-example, consider the transition operator $T \in \mathcal{L}(\mathbb{R}^3)$ given by the stochastic matrix under the standard basis

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}$$

The matrix is clearly stochastic because the entries are non-negative and the columns sum to one. However, note that not all entries are strictly positive, as we have zeroes. Thus, the Perron-Frobenius theorem does not apply. The eigenvalues of T are 1 and 0.5. The eigenvectors associated with the eigenvalue 1 are $\text{span}(e_1, e_2)$ and the eigenvectors associated with the eigenvalue 0.5 are $\text{span}((-1, 0, 1))$. The steady state of T are eigenvectors associated with eigenvalue 1, and they are clearly not unique, as I could construct more than one (for instance, we could construct steady states $(1, 0, 0)$, $(0, 1, 0)$, $(0.5, 0.5, 0)$, and so on. What this implies is that different initial states will result in different steady states. Let's see what happens from $t = 0$ to $t = 50$ for a starting state $x_0 = (0.7, 0.2, 0.1)$ (the red, blue, and green lines represent the first, second, and third entries in the state vectors, respectively):



The system seems to approach a steady $(0.8, 0.2, 0)$. Here is what happens to $x_0 = (0.1, 0.7, 0.2)$ from $t = 0$ to $t = 50$:



From a different starting state, the system approaches a steady state $(0.7, 0.3, 0)$.

The Perron-Frobenius theorem and theorem 2 allow us to quickly identify if a Markov chain will converge to a unique steady state based strictly on the structure of its corresponding transition operator. This is a very useful finding for any practical application!

5.3 Discussion

In this section, I explored Markov Chains through a practical perspective, solving a problem relating to modeling population movement. I also explored the nature of the steady

state through a more rigorous approach. Through my exploration, I demonstrated that the structure of any Markov Chain's transition operator can quickly reveal the nature of convergence, with proofs wielding concepts from the matrices of linear operators, the adjoints of operators, and eigenvectors and eigenvalues.

References

- [1] S. Axler, *Linear Algebra Done Right*, 3, Ed. Springer, 2017.
- [2] D. Margalit and J. Rabinoff, *Interactive Linear Algebra*. Georgia Institute of Technology, 06 2019.
- [3] O. Knill, “Linear algebra with probability notes,” Spring 2011.

A Additional Proofs and Results

A.1 Proof that Differentiation on Polynomials is a Linear Map

Suppose D is the function which takes polynomials in $\mathcal{P}_n(\mathbb{R})$ to $\mathcal{P}_{n-1}(\mathbb{R})$ for integers $n \geq 1$ defined by $Dp = p'$ for all $p \in \mathcal{P}_n(\mathbb{R})$. I claim this is a linear map.

Proof. Suppose $p, q \in \mathcal{P}_n(\mathbb{R})$. Then, $D(p + q) = (p + q)' = p' + q' \in \mathcal{P}_{n-1}(\mathbb{R})$, because the highest degree term in p and q will have its degree decremented by one. The highest possible degree in $p + q$ is n , so $p' + q'$ will have maximal degree $n - 1$. The degree of the polynomial doesn't change under addition. Next, for $\lambda \in \mathbb{R}$, $D(\lambda p) = (\lambda p)' = \lambda p' \in \mathcal{P}_{n-1}(\mathbb{R})$, because multiplication by a constant does not alter the degree of a polynomial. Finally, $D(0) = 0' = 0 \in \mathcal{P}_{n-1}(\mathbb{R})$. With these three checks, D is a linear map from $\mathcal{P}_n(\mathbb{R})$ to $\mathcal{P}_{n-1}(\mathbb{R})$. That is $D \in \mathcal{L}(\mathcal{P}_n(\mathbb{R}), \mathcal{P}_{n-1}(\mathbb{R}))$. \square

A.2 Inner Product from Section 3, Problem 3

We introduced the inner product on $\mathcal{P}_3(\mathbb{R})$ by

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx,$$

for all $f, g \in \mathcal{P}_3(\mathbb{R})$. Here is a proof of why this is an inner product:

Proof. An inner product must satisfy the following four conditions:

1. Positivity

We want to show that $\langle f, f \rangle \geq 0$ for any $f \in \mathcal{P}_3(\mathbb{R})$. For any such f ,

$$\langle f, f \rangle = \int_0^1 (f(x))^2 dx.$$

Clearly, $f(x)^2 \geq 0$ will always be true as we are dealing with real polynomials, so the area given by the integral will also be non-negative.

2. Definitiveness

We want to show that $\langle f, f \rangle = 0$ if and only if $f = 0$. This is clearly true because

$$\langle f, f \rangle = \int_0^1 (f(x))^2 dx,$$

and $(f(x))^2 \geq 0$ for all f . Any $f \neq 0$ will result in $(f(x))^2 > 0$, and thus the integral will also be strictly positive. Only $f = 0$ will result in $\langle f, f \rangle = 0$. If $f = 0$, then

$$\langle f, f \rangle = \langle 0, 0 \rangle = \int_0^1 0^2 dx = 0,$$

as desired.

3. Additivity in first slot

We want to show that for $f, g, h \in \mathcal{P}_3(\mathbb{R})$, $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$. This is true because

$$\langle f + g, h \rangle = \int_0^1 (f(x) + g(x))h(x)dx = \int_0^1 f(x)h(x) + g(x)h(x)dx$$

and

$$\int_0^1 f(x)h(x)dx + \int_0^1 g(x)h(x)dx = \langle f, h \rangle + \langle g, h \rangle,$$

as desired.

4. Homogeneity in first slot

We want to show that for $f, g \in \mathcal{P}_3(\mathbb{R})$ and $\lambda \in \mathbb{R}$, $\langle \lambda f, g \rangle = \lambda \langle f, g \rangle$. This is true because

$$\langle \lambda f, g \rangle = \int_0^1 \lambda f(x)g(x)dx = \lambda \int_0^1 f(x)g(x)dx = \lambda \langle f, g \rangle,$$

as desired.

5. Conjugate Symmetry

We want to show that $\langle f, g \rangle = \overline{\langle g, f \rangle}$ for any $f, g \in \mathcal{P}_3(\mathbb{R})$. We observe that $f(x), g(x) \in \mathbb{R}$ for any $x \in \mathbb{R}$, and so

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx \in \mathbb{R}.$$

So,

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx = \int_0^1 g(x)f(x)dx = \langle g, f \rangle = \overline{\langle g, f \rangle},$$

as desired. This is true because the conjugate of any real number is itself.

Since all four inner product conditions are satisfied, we completed the proof. \square

A.3 Eigenvalues of Normal Operators

In section 5.2.2 we relied on the fact for some transition operator T , if 1 is an eigenvalue of T^* , then it is also an eigenvalue of T . The more generalized form of this result is that for any finite-dimensional inner product space V and any $T \in \mathcal{L}(V)$, if $\lambda \in \mathbb{F}$ is an eigenvalue of T^* , then $\bar{\lambda}$ is an eigenvalue of T . We proceed with a direct proof.

Proof. Let $\lambda \in \mathbb{F}$. Then λ is not an eigenvalue of T^* if and only if $T^* - \lambda I$ is invertible. If this is the case, then there exists some operator $S \in \mathcal{L}(V)$ such that $S(T^* - \lambda I) = (T^* - \lambda I)S = I$. If we take the adjoint on each side of that equality, we get $(T - \bar{\lambda}I)S^* = S^*(T - \bar{\lambda}I) = I$. Thus, we get that $(T^* - \lambda I)$ is invertible if and only if $(T - \bar{\lambda}I)$ is invertible, and therefore $\bar{\lambda}$ is an eigenvalue of T if and only if λ is an eigenvalue of T^* . This completes the proof. \square

Therefore, if λ is an eigenvalue of T^* , then $\bar{\lambda}$ is an eigenvalue of T . In the particular case of stochastic operators as in section 5.2.2, if we set $\lambda = 1$ to be an eigenvalue of T^* , the adjoint of a stochastic operator $T \in \mathcal{L}(\mathbb{R}^n)$, then $\bar{\lambda} = 1$ is also an eigenvalue of T .

The opposite direction of the statement is also true. That is, if λ is an eigenvalue of T , then $\bar{\lambda}$ is an eigenvalue of T^* . The proof is the same, except we swap occurrences of T^* with T .