Abstract Linear Algebra

Tim Smits

May 26, 2023

Contents

Introduction

These notes arose from a math 115A course I was a teaching assistant for at UCLA in fall 2018. They contain, in my opinion, what all students should learn in a first course on abstract vector spaces. At UCLA, math 115A is taken by students from various disciplines. Many students have a hard time adjusting to the abstract nature of the course. These notes try to provide many abstract examples to make the transition easier, as well as provide some interesting, non-standard examples of how the language of linear algebra can be applied to solve (hopefully) interesting problems. There may be many typos. Let me know if any are found!

Chapter 1

Vector Spaces

Linear algebra arose out of trying to solve systems of linear equations. In a first course, one learns the proper way to think about solutions to a system of n linear equations in n variables is by viewing them as solutions to a certain matrix equation of the form $Ax = b$ in $\mathbb{R}^n$, and then develops the necessary theory of matrices and tools needed to solve these equations. The goal of abstract linear algebra is to capture the special properties of $\mathbb{R}^n$ and of matrices that made the theory useful in the first place, and expand it to work in larger settings. This will lead to the abstract definitions of vector spaces and linear transformations, which will be the main objects of study for us.

1.1 Basic Definitions

Definition 1.1.1. A **vector space** V over a field F is a set V with an addition operation $+$ and scalar multiplication operation $\cdot$ by elements of F that sasify the following axioms:

1. For all $x, y \in V$, $x + y = y + x$.
2. For all $x, y, z \in V$, $(x + y) + z = x + (y + z)$.
3. There exists $0 \in V$ such that $x + 0 = x$ for all $x \in V$.
4. For all $x \in V$, there exists $-x \in V$ such that $x + (-x) = 0$.
5. For all $x \in V$, $1 \cdot x = x$.
6. For all $a, b \in F$ and $x \in V$, $(a + b) \cdot x = a \cdot x + b \cdot x$.
7. For all $a, b \in F$ and $x \in V$, $(ab) \cdot x = a \cdot (b \cdot x)$.
8. For all $a \in F$ and $x, y \in V$, $a \cdot (x + y) = a \cdot x + a \cdot y$.

The elements of V are called **vectors**, and it is understood that we write cx to mean $c \cdot x$.

From the axioms above, one can deduce the usual algebraic rules are true in vector spaces:

Proposition 1. *Let V be a vector space. For $x, y, z \in V$, and $a \in F$, the following hold:*

1. If $x + y = x + z$, then $y = z$.

2. The vectors 0 and $-x$ are unique.

3. $0 \cdot x = a \cdot 0 = 0$.

4. $(-a)x = -(ax)$.

The proofs of the above all follow quickly from the vector space and field axioms, and are left as an exercise.

Definition 1.1.2. For a vector space V and $W \subset V$, we call W a **subspace** of V if W is vector space under the same operations as in V.

Proposition 2 (Subspace criterion)**.** *Let V be a vector space. Then $W \subset V$ is a subspace $\iff$ $0 \in W$, and W is closed under addition and scalar multiplication.*

Proof. The forward direction is immediate by definition of a vector space. Conversely, if W is closed under addition and scalar multiplication, since vectors in W are vectors in V, this immediately gives axioms $1, 2, 5, 6, 7, 8$. Since $0 \in W$, 3 is satisfied, and since W is closed under scalar multiplication $(-1)x = -x \in W$ so 4 is satisfied. □

Example 1.1.3. Any field F is a vector space over itself. More generally, F^n is an F-vector space for any n with operations of addition and scalar multiplication performed componentwise, where $F^n = \{(a_1, \dots, a_n) : a_i \in F\}$ is the set of all n-tuples with entries in F.

Example 1.1.4. Let $F \subset L$ be a subfield. Then L is an L-vector space because L is a field, but L is also an F-vector space, with scalar multiplication by an element of F given by performing the multiplication in L, and addition also performed in L.

Example 1.1.5. $M_n(F)$, the set of $n \times n$ matrices with entries in a field F, is an F-vector space, with addition and scalar multiplication done entrywise.

Example 1.1.6. Let S be a non-empty set, then the set of all functions from S to F, denoted $\mathcal{F}(S, F)$, is an F-vector space. A vector in $\mathcal{F}(S, F)$ is a function $f : S \to F$, and two vectors f, g are equal if $f(s) = g(s)$ for all $s \in S$. The operations are given by $(f+g)(s) = f(s)+g(s)$ and $(cf)(s) = cf(s)$.

Example 1.1.7. The set of polynomials of degree at most n with coefficients in F, $P_n(F)$, is a subspace of $\mathcal{F}(F, F)$.

Example 1.1.8. Let $C([a, b])$ be the set of continuous functions from $[a, b]$ to $\mathbb{R}$, and let $C^\infty([a, b])$ be the subset of all infinitely differentiable functions. Then both $C([a, b])$ and $C^\infty([a, b])$ are subspaces of $\mathcal{F}(\mathbb{R}, \mathbb{R})$, and $C^\infty([a, b])$ is a subspace of $C([a, b])$. Let $V = \{f \in C^\infty([a, b]) : f' = f\}$. Then V is a subspace of $C^\infty([a, b])$. This gives a connection between studying solutions to differential equations and studying the vector space $C^\infty([a, b])$.

Example 1.1.9. Let S be a set and set $V = 2^S$, the set of all subsets of S. Then V is a vector space over $\mathbb{Z}/2\mathbb{Z}$ with addition given by $A+B = (A\backslash B)\cup(B\backslash A)$, scalar multiplication in the obvious manner, and the 0 element being the empty set. Note the additive inverse of any set A is itself.

1.2 Operations on Vector Spaces

A natural question to ask is what operations can we do on vector spaces to create new vector spaces? Below are some examples:

Proposition 3. *Let V and W be vector spaces. Then $V \cap W$ is a subspace of both V and W.*

Proof. Clearly $0 \in W \cap V$. If $x, y \in V \cap W$, then as both V, W are vector spaces $x + y$ lies in both V and W, so $x + y \in V \cap W$, and similarly for $c \in F$ we see $cx \in V \cap W$. □

Proposition 4. *Let V and W be vector spaces. Then $V \cup W$ is a vector space $\iff$ $W \subset V$ or $V \subset W$.*

Proof. The backwards direct is immediate: if $W \subset V$ or $V \subset W$, then the union is equal to either V or W which is a vector space. Conversely, suppose that $V \cup W$ is a subspace and that $W \not\subset V$. Suppose for contradiction that $V \not\subset W$. Pick $w \in W \setminus V$, and $v \in V \setminus W$. Then both $w, v \in W \cup V$, so $w + v \in W \cup V$. If $w + v \in W$, then $(w + v) - w = v \in W$, a contradiction. Similarly, if $w + v \in V$ then $w \in V$, again impossible. Therefore $V \subset W$. A similar argument shows that if $V \not\subset W$ then $W \subset V$. □

The above proposition says that taking unions of vector spaces won't produce anything new. However, there is a way to create a vector space that contains copies of both V and W:

Proposition 5. *Let V and W be vector spaces. Then $V \times W$ is a vector space with the operations of componentwise addition and scalar multiplication.*

Proof. Exercise. □

The vector space $V \times W$ is sometimes called the **external direct sum** of V and W and is commonly denoted $V \oplus W$. However to avoid confusion with the definition below, we'll keep the notation $V \times W$. There is a way to "add" vector spaces, but only if they are both subspaces of some common vector space, so that addition of vectors makes sense.

Definition 1.2.1. Let W_1 and W_2 be subspace of a vector space V. The **sum** of W_1 and W_2, denoted $W_1 + W_2$ is defined as $W_1 + W_2 = \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\}$. Further, if $W_1 \cap W_2 = \{0\}$, then we call $W_1 + W_2$ the **interal direct sum** of W_1 and W_2 and denote this $W_1 \oplus W_2$.

The difference between external and interal direct sums is that in the latter case, both spaces live internally inside a larger vector space to begin with. In an external direct sum, we create a larger vector space in which copies of V and W can be identified, namely we identify V with the subspaces $\{(v, 0) : v \in V\} = V \times \{0\}$ and W with $\{(0, w) : w \in W\} = \{0\} \times W$. A sum being internal or external is to be understood by the context, and will just be referred to as a direct sum.

Proposition 6. *For subspaces W_1, W_2 of a vector space V, $W_1 + W_2$ (and therefore $W_1 \oplus W_2$) is a subspace of V.*

Proof. It's clear that $0 \in W_1 + W_2$ since $0 \in W_1$ and $0 \in W_2$. If $x, y \in W_1 + W_2$, then $x = w_1 + w_2$ and $y = w_1' + w_2'$ for $w_1, w_1' \in W_1$ and $w_2, w_2' \in W_2$. Therefore $x + y = (w_1 + w_1') + (w_2 + w_2')$ and $w_1 + w_1' \in W_1'$ because W_1 is a subspace of V, and $w_2 + w_2' \in W_2$ for the same reasoning. The proof that $W_1 + W_2$ is closed under scalar multiplication is similar. □

The difference between being a sum of subspaces and a direct sum of subspaces is the following:

Proposition 7. *Suppose $V = W_1 + W_2$ for some subspaces W_1, W_2. Then $V = W_1 \oplus W_2 \iff$ every vector x in V can be written uniquely as $x = w_1 + w_2$ for $w_1 \in W_1$ and $w_2 \in W_2$.*

Proof. If $V = W_1 \oplus W_2$, and x has two different representations as a sum of the above form, write $x = w_1 + w_2$ and $x = w_1' + w_2'$ for some $w_1, w_2, w_1', w_2' \in W$. Then $w_1 - w_1' = w_2' - w_2$, and the left hand side lives in W_1 while the right hand side lives in W_2. This says $w_1 - w_1' \in W_1 \cap W_2 = \{0\}$, so $w_1 = w_1'$. Similarly $w_2 = w_2'$ so the representation is unique. Conversely, suppose that any vector $x \in V$ can be written uniquely as $x = w_1 + w_2$ for some $w_1 \in W_1$ and $w_2 \in W_2$. Then clearly, $V = W_1 + W_2$. If $x \in W_1 \cap W_2$, we can write $x = x + 0$ by taking $w_1 = x$ and $w_2 = 0$. Similarly, we can write $x = 0 + x$ by taking $w_1 = 0$ and $w_2 = x$. By uniqueness, this says $x = 0$, so that $W_1 \cap W_2 = \{0\}$ says $V = W_1 \oplus W_2$. □

Example 1.2.2. In $\mathbb{R}^2$, set $X = \{(x, y) : y = 0\}$ and $Y = \{(x, y) : x = 0\}$. Then $\mathbb{R}^2 = X \oplus Y$. Note these subspaces are simply the x and y axes. In $\mathbb{R}^3$, set $V = \{(x, y, z) : z = 0\}$ and $W = \{(x, y, z) : x = 0\}$. Then $\mathbb{R}^3 = V + W$, but the sum is not direct because $V \cap W = \{(x, y, z) : x = z = 0\}$.

Example 1.2.3. Let F be a field not of characteristic 2, and let $\mathrm{Sym}_n(F), \mathrm{Skew}_n(F) \subset M_n(F)$ be the subspaces of symmetric and skew-symmetric matrices respectively. Then $M_n(F) = \mathrm{Sym}_n(F) \oplus \mathrm{Skew}_n(F)$. Any matrix $A \in M_n(F)$ can be written $A = \frac{1}{2}(A + A^t) + \frac{1}{2}(A - A^t)$, so $M_n(F) = \mathrm{Sym}_n(F) + \mathrm{Skew}_n(F)$, and if $A \in \mathrm{Sym}_n(F) \cap \mathrm{Skew}_n(F)$, we have $A = A^t$ and $A = -A^t$ so that $2A^t = 0$ says $A^t = 0$, so that the sum is direct.

There's one more common operation on subspaces that we'll study, although it is quite a bit more abstract.

Definition 1.2.4. Let V be a vector space, and $W \subset V$ be a subspace. For a vector $v \in V$, we define the **coset** of v, denoted $v + W$, to be $v + W = \{v + w : w \in W\}$, the set of translates of v by elements of W.

Example 1.2.5. Let $V = \mathbb{R}^2$, $W = \{(x, 0) : x \in \mathbb{R}\}$, and $v = (0, 1)$. What set is $v + W$? Elements of the coset $v + W$ look like $(0, 1) + w$ for different choices of vectors $w \in W$. Since an arbitrary $w \in W$ looks like $(a, 0)$ for some $a \in \mathbb{R}$, such elements look like $(a, 1)$ for some $a \in \mathbb{R}$. For any choice of a the vector $(a, 0)$ is in W, so we see that $v + W = \{(a, 1) : a \in \mathbb{R}\}$.

The point of cosets is that they give us a way of partitioning the vector space V: as an equality of sets, we have $V = \bigcup\limits_{v \in V} (v + W)$. We'll use these cosets to construct a new vector space. Let $V/W = \{v + W : v \in V\}$. We can define addition and scalar multiplication operations on V/W as follows:

Proposition 8. *V/W is a vector space, where the operations are given by $(v+W)+(v'+W) = (v+v')+W$ and $c \cdot (v+W) = c \cdot v + W$.*

Proof. Exercise. □

Definition 1.2.6. The set V/W with the operations of addition and scalar multiplication as given above is known as the **quotient space** of V by W.

The idea behind the quotient space is that it "crushes" the subspace W to the 0 vector. This can be seen from the following:

Proposition 9. *Two cosets $v+W$ and $v'+W$ are equal in V/W if and only if $v - v' \in W$. In particular, $v+W = 0+W$ in V/W if and only if $v \in W$.*

Proof. Exercise. □

Example 1.2.7. Consider $V = \mathbb{R}^2$ and $W = \{(x, 0) : x \in \mathbb{R}\}$, the x-axis. For any vector $v = (a, b)$, we have that $v + W = \{(a+x, b) : x \in \mathbb{R}\}$ is the horizontal line through the vector v. The quotient space V/W "crushes" each of these horizontal lines to a single point, namely the intersection of this line with the y-axis: in the quotient space, we have the equality $(a,b)+W = (0,b)+W$ because $(a,b)-(0,b) = (a,0) \in W$. We see that points in V/W can be "identified" with points on the y-axis, so that one can "picture" V/W as the y-axis.

1.3 Linear Independence

The above discussion tells us how to create new vector spaces from subspaces of some V. How can we create subspaces of V? Starting with $S \subset V$, what is needed to build a vector space out of elements of S? By definition, such a subspace would have to be closed under scalar multiplication, so for $s \in S$ and $c \in F$ it must contain $c \cdot s$. Similarly, it would need to be closed under addition, so it needs to contain all possible finite sums of the elements of the form just mentioned. It turns out, this is enough.

Definition 1.3.1. A **linear combination** of vectors $v_1, \ldots, v_n$ is an expression of the form $c_1v_1 + \ldots + c_nv_n$ for some $c_i \in F$. An equation of the form $c_1v_1 + \ldots + c_nv_n = 0$ is called a **linear dependence relation**. A dependence relation is called **trivial** if the only possible solution is when all $c_i = 0$, and is called non-trivial otherwise.

Definition 1.3.2. Let $S \subset V$. The **span** of S denoted $\text{Span}(S)$ is the set of all finite linear combinations of elements of S. That is, $\text{Span}(S) = \{c_1v_1 + \ldots + c_nv_n : c_i \in F, v_i \in S, n \geq 1\}$.

Proposition 10. *Let $S \subset V$. Then Span(S) is a subspace of V.*

Proof. By convention, if $S = \emptyset$ we define $\text{Span}(S) = \{0\}$. If $S \neq \emptyset$, pick $v \in S$. Then $0 = 0 \cdot v \in \text{Span}(S)$. If $x, y \in \text{Span}(S)$, then $x = c_1v_1 + \ldots + c_nv_n$ and $y = d_1w_1 + \ldots + d_mw_m$ for some $c_i, d_j \in F$ and $v_i, w_j \in S$. Then $x + y = c_1v_1 + \ldots + c_nv_n + d_1w_1 + \ldots + d_mw_m$ is a linear combination of the vectors $v_1, \ldots, v_n, w_1, \ldots, w_m \in S$ so $x + y \in \text{Span}(S)$. Similarly for $c \in F$ we see $c \cdot v \in \text{Span}(S)$, so $\text{Span}(S)$ is a subspace of V. □

Span(S) is sometimes referred to as the subspace generated by S. If $V = \text{Span}(S)$, then we call S a **generating set** for V. Observe that any subspace of V containing S must contain Span(S), and therefore Span(S) is the *smallest* subspace of V containing S.

Example 1.3.3. In $\text{Sym}_2(F)$, any symmetric matrix is of the form $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ for some $a, b, c \in F$. We see $\text{Sym}_2(F) = \text{Span}\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$.

Example 1.3.4. Let $V = \mathbb{R}^3$. Then $V = \text{Span}\{(1,0,0),(0,1,0),(0,0,1)\}$. We may also write $V = \text{Span}\{(1,1,0),(0,1,1),(1,0,1)\})$, or $V = \text{Span}\{(1,1,1),(1,0,0),(0,1,0),(0,1,1)\}$. A spanning set need not be unique, nor must any spanning set have the same cardinality.

The above example shows that a spanning set may contain "redundant" information. In the third spanning set above, notice that $(1,1,1)$ is already contained in $\text{Span}\{(1,0,0),(0,1,0),(0,1,1)\}$, so removing it from S does not change Span(S). We give this condition a name:

Definition 1.3.5. Let V be a vector space. For $S \subset V$, we call S **linearly dependent** if there exist $v_1, \ldots, v_n \in S$ and $c_1, \ldots, c_n \in F$ not all 0 such that $c_1v_1 + \ldots + c_nv_n = 0$. S is called **linearly independent** if S is not linearly dependent.

If S is linearly dependent, the above says there is a non-trivial linear combination of some vectors in S that equals 0. Since some coefficient c_i must be non-zero, we can solve for v_i in terms of the remaining vectors, so another way of saying this is that some vector v_i is contained in the span of some other vectors.

Example 1.3.6. In the above example, the set $\{(1,0,0),(0,1,0),(0,0,1)\}$ is linearly independent. The set $\{(1,1,1),(1,0,0),(0,1,0),(0,1,1)\}$ is linearly dependent.

Example 1.3.7. In $C^\infty(\mathbb{R})$, the vectors $\sin(x)$ and $\cos(x)$ are linearly independent: if $c_1 \sin(x) + c_2 \cos(x) = 0$ for all x, plugging in $x = 0$ and $x = \frac{\pi}{2}$ shows that $c_1 = c_2 = 0$. Similarly if $r \neq s$, the functions e^{rx} and e^{sx} are linearly independent.

Since linear dependence is defined in terms of a finite quantity, an easy definition of linear independence that handles the case of S being infinite is as follows:

Proposition 11. *Let V be a vector space and $S \subset V$. Then S is linearly independent if and only if all finite subsets of S are linearly independent.*

Proof. If S is linearly independent, for any $S' \subset S$, a linear dependence relation among vectors in S' is also a linear dependence relation among vectors in S, so it must be trivial. Conversely, if all finite subsets of S are linearly independent, given any vectors $v_1, \ldots, v_n$, if $c_1v_1 + \ldots + c_nv_n = 0$, this is a dependence relation among the vectors of the finite set $S' = \{v_1, \ldots, v_n\}$, and so must be trivial by assumption. □

Given $S \subset V$, how can we check if S is linearly independent? One way is as follows:

Proposition 12. *Let V be a vector space and $S = \{v_1, \dots, v_n\}$ for some $v_i \in V$. Then S is linearly dependent if and only if $v_1 = 0$ or there exists $1 \leq k < n$ such that $v_{k+1} \in$ Span$(\{v_1, \dots, v_k\})$.*

Proof. The backwards direction is immediate, so suppose that S is linearly dependent. Then $c_1 v_1 + \ldots + c_n v_n = 0$ for some c_i not all 0. Set $k = \max\{n : c_n \neq 0\}$, which exists since some coefficient is non-zero and there are finitely many. Notice that this says $c_i = 0$ for all $k < i \leq n$. If $k = 1$, this says c_1 is the only non-zero coefficient, so $c_1 v_1 = 0$ gives $v_1 = 0$. Otherwise, $k > 1$ so $c_1 v_1 + \ldots + c_n v_n = c_1 v_1 + \ldots + c_k v_k = 0$. Since $c_k \neq 0$, this says $v_k \in \text{Span}(\{v_1, \dots, v_{k-1}\})$, so we are done. □

This gives a method of checking if a set is linearly independent that works well for sets of small size. For example, to check if $\{v_1, v_2, v_3\}$ is linearly independent one just needs to check that $v_2 \notin \text{Span}(\{v_1\})$ and $v_3 \notin \text{Span}(\{v_1, v_2\})$. For sets of larger size, we will later develop more efficient methods. We end the section with an extremely useful proposition.

Proposition 13. *Let $S \subset V$ be linearly independent and $v \in V$. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in$ Span(S).*

Proof. If $S \cup \{v\}$ is linearly dependent, then there are $s_1, \dots, s_n \in S$ and $c_1, \dots, c_{n+1} \in F$ not all 0 such that $c_1 s_1 + \ldots + c_n s_n + c_{n+1} v = 0$. Necessarily, $c_{n+1} \neq 0$ otherwise the linear independence of S forces all $c_i = 0$. Then solving for v gives $v = -\frac{1}{c_{n+1}}(c_1 s_1 + \ldots + c_n s_n)$ so $v \in \text{Span}(S)$. Conversely, if $v \in \text{Span}(S)$ then $v = c_1 s_1 + \ldots + c_n s_n$ for some $s_i \in S$ and $c_i \neq 0$. Then $c_1 s_1 + \ldots + c_n s_n - v = 0$ is a non-trivial linear dependence relation among elements of $S \cup \{v\}$, so $S \cup \{v\}$ is linearly dependent. □

Theorem 1.3.8. *Let $S \subset V$ be linearly independent. If $v \neq 0 \in$ Span(S), then $v = c_1 v_1 + \ldots + c_n v_n$ for unique distinct vectors $v_i \in S$ and unique $c_i \neq 0 \in F$.*

Proof. Suppose that v has two different representations using vectors in S. Write $v = c_1 s_1 + \ldots + c_n s_n$ and $v = d_1 t_1 + \ldots + d_m t_m$ for some $c_i, d_j \neq 0 \in F$ and $s_i, t_j \in S$, where we may assume none of the s_i are the same and none of the t_j are the same. Subtracting shows $c_1 s_1 + \ldots + c_n s_n - d_1 t_1 - \ldots - d_m t_m = 0$. If $\{s_1, \dots, s_n\} \neq \{t_1, \dots, t_m\}$, then there is some i such that s_i is not equal to any of t_j. Since S is linearly independent, this forces $c_i = 0$, since there is no other term in the sum that can be grouped with $c_i s_i$. This is a contradiction, so $n = m$ and $\{s_1, \dots, s_n\} = \{t_1, \dots, t_m\}$. Relabeling as necessary, we may assume that $s_i = t_i$ so that the above can be written as $(c_1 - d_1)s_1 + \ldots + (c_n - d_n)s_n = 0$, so $c_i = d_i$ for all i and therefore such a representation is unique. □

1.4 Bases and Dimension

The above theorem is of critical importance: the vectors in Span(S) can then be though of as tuples of elements of F by reading off the coefficients in the corresponding linear combination. A natural question to ask is if every vector space arises as the spanning set of a linearly independent subset. The answer is yes, and is the most important result in linear algebra.

Definition 1.4.1. A **basis** of a vector space V is a linearly independent spanning set. The **dimension** of V is the cardinality of a basis of V.

Perhaps in more familiar terms, the above says that every vector space has a basis. The fact that the dimension of a vector space is actually well defined is a fairly non-trivial result, but the proof is a rather technical set theoretic argument that is unenlightening, so for our purposes it will be taken for granted.

Proposition 14. *Let B and B' be two bases of a vector space V. Then $|B| = |B'|$.*

Dimension is one of the most useful ideas in linear algebra: it gives us a notion of size for a vector space, and being able to translate questions about vector spaces into statements about integers makes them easier to understand. At this stage, linear algebra branches off in two directions: the study of infinite dimensional vector space, and the study of finite dimensional vector spaces, the latter of which we will focus the majority of our attention on.

1.5 Finite Dimensional Vector Spaces

Throughout the rest of this section, we will assume that V is an n-dimensional vector space over a field F unless otherwise stated.

The above proof that every vector space has a basis is non-constructive – it tells us one must exist but gives us no way of finding one. In the finite dimensional case, we actually have a constructive method for finding bases of a vector space.

Theorem 1.5.1. *Let $S = \{v_1, \ldots, v_k\}$ be a subset of V that spans V. Then there is $B \subset S$ such that B is a basis of V.*

Proof. We may assume that the v_i are non-zero, otherwise remove them. Let m be the largest integer such that there is an m element subset B of S that is linearly independent. As $\{v_i\}$ is linearly independent for any i, and S has at most k elements, in particular B must exist and $1 \leq m \leq k$. Then $\text{Span}(S) = \text{Span}(B)$. To see this, we show that $v_i \in \text{Span}(B)$ for all i. If $v_i \notin B$, then $B \cup \{v_i\}$ is a linearly dependent subset by definition of B, so there are $c_1, \ldots, c_{m+1} \in F$ not all 0 such that $c_1 s_1 + \ldots + c_m s_m + c_{m+1} v_i = 0$. By linear independence of elements of B, necessarily $c_{m+1} \neq 0$, so we can solve for v_i in terms of s_i, giving $v_i \in \text{Span}(B)$ as desired. □

Theorem 1.5.2. *Let $S = \{v_1, \ldots v_k\}$ be a linearly independent subset of V. Then there exist vectors $w_1, \ldots, w_m \in V$ such that $\{v_1, \ldots, v_k, w_1, \ldots, w_m\}$ is a basis of V.*

Proof. Pick a basis $\{e_1, e_2, \ldots, e_n\}$ of V. Then $\{w_1, \ldots, w_k, e_1, \ldots, e_n\}$ is a spanning set. Remove vectors e_i from the above set if $e_i \in \text{Span}(S)$. The remaining vectors $w_1, \ldots, w_m$ not removed are not contained in $\text{Span}(S)$, so the set $\{v_1, \ldots, v_k, w_1, \ldots, w_m\}$ must be linearly independent, and it remains a spanning set of V by construction so it is a basis. □

An immediately corollary is the following:

Corollary 1.5.3. *Let $W \subset V$ be a subspace. Then there exists $W' \subset V$ a subspace such that $V = W \oplus W'$.*

Proof. Pick a basis $\{v_1, \ldots, v_k\}$ of W and extend to a basis $\{v_1, \ldots, v_k, e_1, \ldots, e_m\}$ of V. Set $W' = \mathrm{Span}(\{e_1, \ldots, e_m\})$. Then it's clear that $V = W + W'$, and $W \cap W' = \{0\}$ because if $x \in W \cap W'$, we can write $x = c_1v_1 + \ldots + c_kv_k$ and $x = d_1e_1 + \ldots + d_me_m$ for $c_i, d_j \in F$, so $c_1v_1 + \ldots + c_kv_k - d_1e_1 - \ldots - d_me_m = 0$ gives all $c_i, d_j = 0$ as these vectors are linearly independent in V. □

The subspace W' is called the complement of W in V.

In linear algebra, it's not uncommon to be interested in finding a basis with some particular choices of basis vectors, so the extension result is quite useful. The following is a translation of the above two results using the language of dimension.

Theorem 1.5.4. *Let $S = \{v_1, \ldots, v_k\}$.*

1. *If S is linearly independent, then $k \leq n$.*
2. *If S spans V, then $k \geq n$*
3. *If $k = n$, S is linearly independent if and only if S is a spanning set.*

Proof. Items 1 and 2 are immediately corollaries of the above two results. To prove 3, If S is linearly independent and S doesn't span V, then there is $v \in V$ such that $S \cup \{v\}$ is linearly independent. But then this says $n+1 \leq n$, a contradiction. Therefore S spans V. Conversely, if S is not linearly independent, we may trim S to a basis B with $n = |B| < |S| = n$, a contradiction. □

Example 1.5.5. In F^n, the vectors e_i where e_i is the vector that is 1 in the i-th coordinate and 0 elsewhere form a basis. It's easy to see that if $c_1e_1 + \ldots c_ne_n = 0$, then $(c_1, \ldots, c_n) = (0, \ldots, 0)$ so $c_i = 0$, and it's obvious this is a spanning set. This is an n-dimensional F-vector space.

Example 1.5.6. In $M_n(F)$, the matrices E_{ij} where E_{ij} is the matrix with (i, j)-th entry equal to 1 and 0 elsewhere is a basis – the argument is the same as above. This is an n^2-dimensional F-vector space.

Example 1.5.7. In $P_n(F)$, the set $\{1, x, \ldots, x^n\}$ is a basis. It's clear that this is a spanning set, so it remains to see linear independence. If $c_0 + c_1x + \ldots c_nx^n = 0$ in $P_n(F)$, then in particular, this holds true for all $x \in F$. The left hand side is a degree at most n polynomial, so it has at most n roots, while the right hand side is 0 everywhere. This is only possible if all coefficients are 0. This is an $n + 1$-dimensional F-vector space.

Example 1.5.8. The space of all polynomials with coefficients in F, $P(F)$ is infinite dimensional: any finite set of polynomials has a maximal degree m, so their F-span is contained in $P_m(F)$. This says no finite subset of $P(F)$ is a spanning set, so it is infinite dimensional as an F-vector space.

Example 1.5.9. In $\mathbb{R}^3$, the set $\{(1,0,1),(1,1,0),(0,1,1)\}$ is a basis, because it is linearly independent: one can check by hand that $(0,1,1) \notin \mathrm{Span}(\{(1,0,1),(1,1,0)\})$.

Example 1.5.10. Let $V = \{(x,y,z) \in \mathbb{R}^3 : x - 2y + z = 0 \text{ and } 2x - 3y + z = 0\}$. Then V is a subspace of $\mathbb{R}^3$, and has basis $\{(1,1,1)\}$.

Example 1.5.11. The dimension of a vector space depends on the underlying field. As a $\mathbb{C}$-vector space, $\mathbb{C}^n$ has dimension n with basis vectors e_j for $1 \leq j \leq n$. However, $\mathbb{C}$ is a 2-dimensional $\mathbb{R}$-vector space: any complex number z is of the form $z = a + bi$ for real a, b, so $\{1, i\}$ is basis. The vectors e_j, ie_j for $1 \leq j \leq n$ form a basis of $\mathbb{C}^n$ as a $2n$-dimensional $\mathbb{R}$-vector space.

Example 1.5.12. For $a \neq 0 \in \mathbb{R}$, the set $\{1, x-a, (x-a)^2, \ldots, (x-a)^n\}$ is a basis for $P_n(\mathbb{R})$: if $c_0 + c_1(x-a) + \ldots + c_n(x-a)^n = 0$ for all x, plugging in $x = a$ shows $c_0 = 0$, and taking derivatives and repeating the argument shows $c_i = 0$. This shows linear independence and since a basis has $n+1$ elements, this is a spanning set. Every polynomial $p(x)$ can be written in the form $p(x) = c_0 + c_1(x-a) + \ldots + c_n(x-a)^n$. One can solve for the coefficients c_i by taking derivatives as necessary and plugging in $x = a$, to see $c_k = \frac{p^{(k)}(a)}{k!}$, recovering the usual Taylor expansion around $x = a$.

To illustrate why dimension is useful, we prove a quick result, which helps us understand a vector space by understanding its subspaces.

Proposition 15. *Let $W \subset V$ be a subspace. Then* $\dim(W) \leq n$. *If* $\dim(W) = n$, *then* $W = V$.

Proof. First we show that W is finite dimensional. If $W = \{0\}$, we are done. Otherwise, pick $w_1 \neq 0 \in W$. If $W = \mathrm{Span}(\{w_1\})$, we are done, otherwise there is $w_2 \in W$ with $w_2 \notin \mathrm{Span}(\{w_1\})$, so $\{w_1, w_2\}$ is linearly independent. Continue choosing vectors $w_3, \ldots, w_k$ in this way such that $\{w_1, \ldots, w_k\}$ is a linearly independent subset of W. Since $W \subset V$, it's also a linearly independent subset of V, so this process must stop before the n-th step, and the termination of this process is equivalent to saying that $W = \mathrm{Span}(\{w_1, \ldots, w_k\})$. This says $\{w_1, \ldots, w_k\}$ is a basis of W, and we have $k \leq n$. If $k = n$, these vectors are actually a basis of V as well, so $W = V$. □

Example 1.5.13. Let $W \subset \mathbb{R}^3$ be a subspace. Then $\dim(W) = 0, 1, 2, 3$. If $\dim(W) = 0$, then $W = \{0\}$, and if $\dim(W) = 3$, then $W = \mathbb{R}^3$. If $\dim(W) = 1$, then $W = \mathrm{Span}(\{v\})$ for some $v \in W$, i.e. W is the line through the origin in the direction of v. If $\dim(W) = 2$, we have $W = \mathrm{Span}(\{v_1, v_2\})$ for some vectors v_1, v_2. Let $v = v_1 \times v_2$, so $x \cdot v = 0$ for all $x \in W$. This defines the equation of a plane with normal vector $v_1 \times v_2$, so that subspaces of $\mathbb{R}^3$ are either $\{0\}$, $\mathbb{R}^3$, lines through the origin or planes through the origin. The dimensions of these objects should hopefully match your own geometric intuition.

Example 1.5.14. Set $V = (\mathbb{Z}/p\mathbb{Z})^2$, which is a 2-dimensional $\mathbb{Z}/p\mathbb{Z}$-vector space with basis vectors $(\bar{1}, \bar{0})$ and $(\bar{0}, \bar{1})$. What are all the subspaces of V? If $W \subset V$ is a subspace, we have $\dim(W) = 0, 1, 2$. If $\dim(W) = 0$ then $W = \{0\}$, and if $\dim(W) = 2$ then $W = V$. If $\dim(W) = 1$, then $W = \mathrm{Span}(\{v\})$ for some non-zero vector v. There are a total of $p^2 - 1$ such vectors v, and each of the $p - 1$ non-zero multiples of v span the same subspace of V. Since the 1-dimensional subspaces of V partition W, we conclude there are $(p^2-1)/(p-1) = p+1$ different 1-dimensional subspaces of V, for a total of $p + 3$.

We end with some useful dimension counting results:

Proposition 16. *Let be V a vector spaces and let W, W_1, W_2 be subspaces.*

(a) $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$.

(b) $\dim(W_1 \oplus W_2) = \dim(W_1) + \dim(W_2)$.

(b) $\dim(V/W) = \dim(V) - \dim(W)$.

Proof. Exercise. □

1.6 Zorn's Lemma

In order to deal with infinite spanning sets, we need some set theory:

Definition 1.6.1. A **partial ordering** on a set S is a binary relation $\leq$ that satisfies the following conditions for all $a, b, c \in P$:

1. (Reflexivity) $a \leq a$.
2. (Anti-symmetry) If $a \leq b$ and $b \leq a$ then $a = b$.
3. (Transitivity) If $a \leq b$ and $b \leq c$ then $a \leq c$.

Definition 1.6.2. A **poset** is a set P with a partial ordering $\leq$. A poset is called **totally ordered** if for all pairs of elements $a, b \in P$, either $a \leq b$ or $b \leq a$. A **chain** C of a poset P is a totally ordered subset of P. For $A \subset P$, an **upper bound** of A is an element m of P such that $x \leq m$ for all $x \in A$.

Example 1.6.3. Let $P = \mathbb{R}$ and let $\leq$ be the usual relation of less than or equal to. Then P is a poset, and P is totally ordered.

Example 1.6.4. Let $P = \mathbb{N}$ and let $a \leq b \iff a \mid b$. This makes P a poset. Under this ordering, we have $3 \leq 6$, but 3 and 5 are not comparable, so P is not totally ordered. The set $P' = \{1, 2, 4, 8, 16\}$ is totally ordered, so it is a chain in P. The element 16 is an upper bound of P'.

Example 1.6.5. Let P a set of subsets of a vector space V, and $\leq$ be **ordering by inclusion**, i.e. $W \leq W' \iff W \subset W'$. Then P is a poset.

The proof that every vector space has a basis is one of many non-constructive existence results in mathematics that follow from *Zorn's lemma*, which is (surprisingly!) equivalent to the Axiom of Choice:

Theorem 1.6.6 (Zorn's lemma)**.** *Let P be a poset such that every chain in P has an upper bound in P. Then P has a maxmimal element with respect to $\leq$. That is, there is an element $m \in P$ such that $x \leq m$ for all $x \in P$.*

We are now ready to prove the theorem:

Theorem 1.6.7. *Every vector space has a basis.*

Proof. Let V be a vector space over some field F. The idea of the proof is as follows: use Zorn's lemma to show that V contains a maximal linearly independent subset B of V (in the sense that there is no linearly independent subset S with $B \subsetneq S$), and then show that B must be a basis of V.

If $V = \{0\}$, then by definition $V = \text{Span}(\emptyset)$, and the empty set is linearly independent. Now suppose that $V \neq \{0\}$ and let $P = \{S \subset V : S \text{ is linearly independent}\}$ be the set of all linearly independent subset of V with an ordering on P given by inclusion. Then $P \neq \emptyset$, because there exists $v \neq 0 \in V$ so $\{v\}$ is a linearly independent subset of V. We now check the conditions of Zorn's lemma. Suppose that $C \subset P$ is a chain, and write $C = \{S_\alpha\}_{\alpha \in I}$ for some indexing set I. Set $M = \bigcup_{\alpha \in I} S_\alpha$. The claim is that M is an upper bound of C that is an element of P. The first statement is immediate by definition: for any $S_\alpha \in C$, we have $S_\alpha \subset \bigcup_{\alpha \in I} S_\alpha$, so $C \leq M$. Therefore, we only need to check that M is a linearly independent subset of V, so that $M \in P$, letting Zorn's lemma kicks in.

Suppose that M is not linearly independent, then there are vectors $s_1, \dots, s_n$ where $s_i \in S_{\alpha_i}$ for some S_{α_i} and scalars $c_1, \dots, c_n \in F$ not all 0 such that $c_1 s_1 + \cdots c_n s_n = 0$. As C is totally ordered, one of the sets $S_{\alpha_1}, \dots, S_{\alpha_n}$ must contain the others, so each of the vectors s_i live in some common set, which we denote S_α. This says there is a non-trivial dependence relation among vectors in S_α, contradicting that S_α is linearly independent (because S_α lives in P!). Therefore, M is a linearly independent subset of V. By Zorn's lemma, P contains a maximal element with respect to inclusion, say B.

To finish up, we need to show that B spans V. Suppose otherwise, then there is some $v \in V$ such that $v \notin \text{Span}(B)$. This says that $B \cup \{v\}$ is a linearly independent subset of V with $B \subset B \cup \{v\}$, contradicting the maximality of B. Therefore B spans V, and we are done. □

It's important to note that the proof only shows that a basis *exists* – it gives absolutely zero indication of what one is. The proof technique of using Zorn's lemma is a rather standard one for proving existence theorems in mathematics (especially in algebra) and is worth understanding.

For vector spaces spanned by finite sets, you saw in lecture that it's not too hard to show that any two bases have the same number of elements. This allows us to define the *dimension* of a vector space. What happens if the vector space has a basis of infinitely many elements? The dimension of a vector space is still well defined, but this now becomes a fairly non-trivial result. Instead of talking about the *number* of elements in a basis, we have to talk about the *cardinality* of the basis, and if you know anything about set theory, there are many different "sizes" of infinite sets which is what causes complications. The proof is a rather technical set theoretic argument that is unenlightening, so we will take it for granted.

Proposition 17. *Let B and B' be two bases of a vector space V. Then $|B| = |B'|$.*

This gives us the following definition that works for any vector space:

Definition 1.6.8. Let V be a vector space. The **dimension** of V is defined as the cardinality of a basis of V. V is said to be **infinite dimensional** if it's dimension is not finite.

If you're familiar with the notion of *countability*, the following are examples of infinite vector spaces of different "sizes":

Example 1.6.9. The vector space $\mathbb{Q}[x]$ is infinite dimensional as a $\mathbb{Q}$-vector space, because the span of any finite set of polynomials has bounded degree. The set $\{1, x, x^2, \ldots\}$ is a basis of $\mathbb{Q}[x]$ as a $\mathbb{Q}$-vector space, and so $\mathbb{Q}[x]$ has countable dimension.

Example 1.6.10. $\mathbb{R}$ is infinite dimensional as a $\mathbb{Q}$-vector space, because any finite dimensional vector space over $\mathbb{Q}$ must be countable, and $\mathbb{R}$ is not countable. It turns out that $\mathbb{R}$ has uncountable dimension as a $\mathbb{Q}$-vector space (but this is much harder to show).

Chapter 2

Linear Transformations

A general philosophy is that to study algebraic structures, one needs to not just study the objects but structure preserving maps between these objects as well. There is no simple explanation for the latter, but historically it has been very productive. When studying how to solve systems of linear equations in $\mathbb{R}^n$, one is naturally led to matrix equations of the form $Ax = b$. A matrix A defines a function $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = Ax$. This function T respects the structure of Euclidean space, in the sense that $T(x + y) = T(x) + T(y)$ and $T(cx) = cT(x)$ for all $x, y \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Since vector spaces are nothing more than abstracted versions of Euclidean space, we should look at abstract analogues of matrices, i.e. functions that preserve the vector space structure.

Unless otherwise stated, through the handout V is a finite dimensional vector space of dimension n over a field F. The letter T will always denote a linear transformation.

2.1 Basic Definitions

Definition 2.1.1. A **linear transformation** $T : V \to W$ between vector spaces V and W over a field F is a function satisfying $T(x + y) = T(x) + T(y)$ and $T(cx) = cT(x)$ for all $x, y \in V$ and $c \in F$. If $V = W$, we sometimes call T a **linear operator** on V.

Note that necessarily a linear transformation satisfies $T(0) = 0$. We also see by induction that for any finite collection of vectors $v_1, \ldots, v_n$ and scalars $c_1, \ldots, c_n \in F$ we have $T(c_1v_1 + \ldots + c_nv_n) = c_1T(v_1) + \ldots + c_nT(v_n)$.

Definition 2.1.2. The **kernel** $\ker(T)$ is defined by $\ker(T) = \{x \in V : T(x) = 0\}$. The **image** $\operatorname{Im}(T)$ is defined by $\operatorname{Im}(T) = \{T(x) : x \in V\}$.

The image and kernel of T are two important subspace of V and W respectively, and we can translate set theoretic statements about injectivity and surjectivity into the language of linear algebra.

Proposition 18. *Let $T : V \to W$ be linear. Then $\ker(T)$ is a subspace of V and $\operatorname{Im}(T)$ is a subspace of W.*

Proof. Since T is linear, we have $T(0) = T(0+0) = T(0) + T(0)$, so $0 = T(0)$ gives $0 \in \ker(T)$. If $x, y \in \ker(T)$ then $T(x+y) = T(x) + T(y) = 0$ by linearity. Similarly, if $c \in F$, $T(cx) = cT(x) = 0$ so $cx \in \ker(T)$ giving $\ker(T)$ is a subspace of V. Since $T(0) = 0$, this says $0 \in \mathrm{Im}(W)$. If $x, y \in \mathrm{Im}(W)$ then there are $u, v \in V$ such that $x = T(u)$ and $y = T(v)$. Then $x + y = T(u) + T(v) = T(u+v)$ so $x + y \in \mathrm{Im}(T)$. Finally, if $x = T(u)$ then $cx = cT(u) = T(cu)$ so $cx \in \mathrm{Im}(T)$ which says $\mathrm{Im}(T)$ is a subspace of W. □

Proposition 19. *Let $T : V \to W$ be linear.*

(a) T is injective if and only if $\ker(T) = \{0\}$.

(b) T is surjective if and only if $\mathrm{Im}(T) = W$.

Proof.

(a) Suppose that T is injective. If $x \in \ker(T)$, then $T(x) = T(0)$ so injectivity says $x = 0$ giving $\ker(T) = \{0\}$. If $\ker(T) = \{0\}$, if $T(x) = T(y)$ then $T(x - y) = 0$ says $x - y \in \ker(T)$ so $x - y = 0$, i.e. $x = y$ so T is injective.

(b) If T is surjective, then for every $y \in W$ there is $x \in V$ such that $T(x) = w$, which is precisely the same as saying $W = \mathrm{Im}(T)$. On the other hand, if $W = \mathrm{Im}(T)$ then for all $w \in W$ there is $x \in V$ with $w = T(x)$, so T is surjective.

□

Example 2.1.3. For any vector space V, the **identity transformation** $\mathrm{id}_V : V \to V$ given by $\mathrm{id}_V(x) = x$ is linear.

Example 2.1.4. For any field F and $a \in F$, the map $T : F \to F$ given by $T(x) = ax$ is a linear transformation by the field axioms.

Example 2.1.5. The map $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(x, y) = (x + y, y, x - y)$ is a linear transformation.

Example 2.1.6. For any matrix $A \in M_{m\times n}(F)$, the map $T : F^n \to F^m$ given by $T(x) = Ax$ is a linear transformation, since $A(x + y) = Ax + Ay$ and $A(cx) = c(Ax)$ by how matrices work.

Example 2.1.7. In $P(\mathbb{R})$, the maps $D(p)(x) = p'(x)$ and $I(p)(x) = \int_0^x p(t)\,dt$ are linear operators on $P(\mathbb{R})$ by calculus. D is not injective, because any constant polynomial has derivative 0, but D is surjective since $D(\int_0^x p(t)\,dt) = p(x)$ by the fundamental theorem of calculus. The operator I is injective but not surjective because nothing maps to the polynomial $p(x) = 1$.

Example 2.1.8. The map $D : C^\infty([0,1]) \to C^\infty([0,1])$ given by $D(f)(x) = f(x) - f'(x)$ is a linear transformation. Saying $f \in \ker(D)$ is the same as saying $f'(x) = f(x)$, so $\ker(D)$ is precisely the set of functions that satisfy the differential equation $f = f'$. From calculus, we know the only such functions are of the form ce^x for $c \in \mathbb{R}$, so $\ker(D) = \mathrm{Span}(\{e^x\})$ is a 1-dimensional subspace of $C^\infty([0,1])$.

Example 2.1.9. The map $T : M_n(F) \to M_n(F)$ given by $T(A) = A - A^t$ is linear. $\ker(T)$ is the set of matrices with $A = A^t$, i.e. $\ker(T) = \mathrm{Sym}_n(F)$. Any matrix in $\mathrm{Im}(T)$ is of the form $A - A^t$ for some A, which is skew-symmetric, so $\mathrm{Im}(T) \subset \mathrm{Skew}_n(F)$. If F does not have characteristic 2, for any skew-symmetric matrix B, we have $T(B) = B - B^t = 2B$, so $T(\frac{1}{2}B) = B$ says $\mathrm{Im}(T) = \mathrm{Skew}_n(F)$.

Example 2.1.10. Let F^∞ be the sequence space of elements of F. That is, $F^\infty = \{(a_1, a_2, \ldots) : a_i \in F\}$. Define maps $R : F^\infty \to F^\infty$ by $R((a_1, a_2, \ldots)) = (a_2, a_3, \ldots)$ and $L((a_1, a_2, \ldots)) = (0, a_1, a_2, \ldots)$, the right and left shift operators respectively. Then both R and L are linear operators on F^∞.

Example 2.1.11. Let $W \subset V$ be a subspace. The map $\pi : V \to V/W$ given by $T(v) = v+W$ is a linear transformation, called the **quotient map**.

Example 2.1.12. Suppose $V = W \oplus U$ for some subspaces W, U of V. The **projection** π_W of V onto W along U is defined by $\pi_W(x) = w$ where $x = w + u$ for unique $w \in W$ and $u \in U$. Then π_W is linear, and $\ker(\pi_W) = U$ and $\mathrm{Im}(\pi_W) = W$. If we assume V is finite dimensional, for any subspace W there is U such that $V = W \oplus U$. This then says that any subspace W is the kernel of some linear transformation, namely π_U where U is the complement of W in V. Similarly, W appears as the image of π_W.

A natural question is given a linear operator $T : V \to V$ and a subspace W of V, when does T restrict to a linear operator on W? Necessarily, if T restricts to an operator on W we must have $T(W) \subset W$, and actually this is sufficient: if $T(W) \subset W$, then for $x, y \in W$ we have $T(x + y) = T(x) + T(y)$ since $x, y \in V$ and $T(cx) = cT(x)$ for $c \in F$. We give such subspaces a name:

Definition 2.1.13. Given a linear operator $T : V \to V$, a subspace $W \subset V$ is called **T-invariant** if $T(W) \subset W$. The restriction of T to W, denoted by $T|_W$, is the linear transformation $T|_W(x) = T(x)$ for all $x \in W$.

Example 2.1.14. Let $V = W \oplus U$, and consider π_W. Then W is π_W-invariant, and $\pi_W|_W$ is the identity map on W.

Example 2.1.15. The map $T : M_n(F) \to M_n(F)$ given by $T(A) = A - A^t$ is $\mathrm{Sym}_n(F)$-invariant. The restriction $T|_{\mathrm{Sym}_n(F)}$ is simply the 0 map.

2.2 Dimension Counting

The image and kernel of a linear transformation T are extremely important because they give rise to a powerful dimension counting result. We'll illustrate how we can use such counting arguments to get useful results.

Definition 2.2.1. The **rank** of a linear transformation $T : V \to W$, denoted by $\mathrm{rank}(T)$ is defined as $\mathrm{rank}(T) = \dim(\mathrm{Im}(T))$.

Proposition 20. *If $\{v_1, \ldots, v_n\}$ is a basis of V, then $\{T(v_1), \ldots, T(v_n)\}$ is a spanning set of $\mathrm{Im}(T)$. If T is injective then $\{T(v_1), \ldots, T(v_n)\}$ is a basis of $\mathrm{Im}(T)$. In particular, if T is injective then rank$(T) = n$.*

Proof. Let $y \in \mathrm{Im}(T)$. Then $y = T(x)$ for some $x \in V$, and we may write $x = c_1v_1 + \ldots + c_nv_n$ for some $c_i \in F$. Then $y = T(x) = T(c_1v_1 + \ldots + c_nv_n) = c_1T(v_1) + \ldots + c_nT(v_n)$, so $y \in \mathrm{Span}(\{T(v_1), \ldots, T(v_n)\}$, which says this is a spanning set of $\mathrm{Im}(T)$. If further we assume that T is injective, if $c_1T(v_1) + \ldots + c_nT(v_n) = 0$, then $T(c_1v_1 + \ldots + c_nv_n) = 0$, so $c_1v_1 + \ldots + c_nv_n \in \ker(T)$. Since T is injective, this says $c_1v_1 + \ldots + c_nv_n = 0$, and since the vectors v_i are linearly independent this says $c_i = 0$, i.e. that $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent and therefore a basis of $\mathrm{Im}(T)$, so that $\mathrm{rank}(T) = n$. □

Theorem 2.2.2. *(Rank-Nullity) Let W be a vector space. If $T : V \to W$ is linear, then rank$(T) + \dim(\ker(T)) = n$.*

Proof. Pick a basis $\{v_1, \ldots, v_k\}$ of $\ker(T)$ and extend this to a basis $\{v_1, \ldots, v_k, w_1, \ldots, w_\ell\}$ of V. The above shows that $\mathrm{Im}(T) = \mathrm{Span}(\{T(w_1), \ldots, T(w_\ell)\})$, so $\{T(w_1), \ldots, T(w_\ell)\}$ is a spanning set and therefore it's sufficient to prove it is a basis of $\mathrm{Im}(T)$. Suppose $c_1T(w_1) + \ldots + c_\ell T(w_\ell) = 0$. Then $T(c_1w_1 + \ldots + c_\ell w_\ell) = 0$, so $c_1w_1 + \ldots + c_\ell w_\ell \in \ker(T)$. We may then write $c_1w_1 + \ldots + c_\ell w_\ell = a_1v_1 + \ldots + a_kv_k$ for some $a_i \in F$, so $c_1w_1 + \ldots + c_\ell w_\ell - a_1v_1 - \ldots - a_kv_k = 0$. Since the vectors w_i, v_j are a basis of V, this says all $c_i = 0$ and all $a_i = 0$, so that $\{T(w_1), \ldots, T(w_\ell)\}$ is a basis as desired. □

Sometimes $\dim(\ker(T))$ is referred to as the nullity of T, hence the name of the theorem, but this terminology is not commonly used outside of linear algebra textbooks. As an immediate corollary of the rank-nullity theorem, we getting the following analogous result for functions between finite sets of the same size:

Corollary 2.2.3. *Let V and W be vector spaces of the same dimension. Then $T : V \to W$ is injective $\iff$ T is surjective $\iff$ T is bijective.*

Proof. T is injective if and only if $\ker(T) = \{0\}$, so by rank-nullity this says $n = \mathrm{rank}(T) + 0$, i.e. $\mathrm{Im}(T) = W$ so T is surjective. Similarly if T is surjective, $\mathrm{rank}(T) = n$ so rank-nullity says $n = n + \dim(\ker(T))$ so $\dim(\ker(T)) = 0$ gives $\ker(T) = \{0\}$ and therefore T is injective. □

We give some examples to illustrate how the rank-nullity theorem is used to compute images and kernels of linear transformations.

Example 2.2.4. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (x+y+2z, 2x+2y+4z, 2x+3y+5z)$. Then T is a linear transformation, and $\mathrm{Im}(T) = \mathrm{Span}(\{(1, 2, 2), (1, 2, 3), (2, 4, 5)\}) = \mathrm{Span}(\{(1, 2, 2), (1, 2, 3)\})$. The latter set is a basis for $\mathrm{Im}(T)$, so that $\mathrm{rank}(T) = 2$, i.e. $\mathrm{Im}(T)$ is a plane in $\mathbb{R}^3$. By rank-nullity the kernel of T is 1-dimensional, so it must be a line. Which line is it? Representing T as the matrix $A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 2 & 3 & 5 \end{pmatrix}$, one sees that any vector orthogonal to the rows of A is contained in the kernel. Taking the cross product of the first and third rows shows $(-1, 1, 1) \in \ker(T)$ so that $\ker(T) = \mathrm{Span}(\{(-1, 1, 1)\})$.

Example 2.2.5. Let $T : M_n(F) \to F$ be the trace map $T(A) = \mathrm{tr}(A)$. Clearly T is surjective, so by rank-nullity we have $\dim(\ker(T)) = n^2 - 1$. For $A \in \ker(T)$, we have $a_{11} + \ldots + a_{nn} = 0$, which says $a_{11} = -a_{22} - \ldots - a_{nn}$. Since the condition on trace has nothing to do with entries off the diagonals, we see that the matrices E_{ij} with $i \neq j$ along the matrices $-E_{11} + E_{ii}$ with $2 \leq i \leq n$ form a spanning set for $\ker(T)$, and therefore a basis because there are $n^2 - 1$ of them.

Example 2.2.6. The map $T : P(\mathbb{R}) \to P(\mathbb{R})$ defined by $T(p) = 5p'' + 3p'$ is surjective. Let q be a polynomial of degree n. Restricting T to $P_{n+1}(\mathbb{R})$ defines a map $T' : P_{n+1}(\mathbb{R}) \to P_n(\mathbb{R})$. with $T'(p) = T(p)$. By rank-nullity, $\text{rank}(T') + \dim(\ker(T')) = n + 2$. If $p \in \ker(T')$, then $5p'' + 3p' = 0$ says $5p'' = -3p'$. Since $\deg(p'') = \deg(p') - 1$, this is impossible unless both p' and p'' are 0, i.e. p is a constant. This says $\ker(T') = \text{Span}(\{1\})$, so $\ker(T')$ is 1-dimensional, and $\text{rank}(T') = n + 1$ says T' is surjective.

Example 2.2.7. Let F not be of characteristic 2, and $T : M_n(F) \to M_n(F)$ be given by $T(A) = A - A^t$. The rank-nullity theorem says that $\dim(\text{Sym}_n(F)) + \dim(\text{Skew}_n(F)) = n^2$. One can check (by explicitly finding a basis) that $\dim(\text{Skew}_n(F)) = \frac{n(n-1)}{2}$, so $\dim(\text{Sym}_n(F)) = n^2 - \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$.

Example 2.2.8. Let $\dim(V) = n$ and $\dim(W) = m$ with $n < m$. Then there is no surjective linear transformation T from V to W: by rank-nullity, $\text{rank}(T) + \dim(\ker(T)) = n$, so $\text{rank}(T) = n - \dim(\ker(T)) \leq n < m$ says $\text{Im}(T) \neq W$. Similarly, if $n > m$ there is no injective linear transformation from V to W. This says if $n < m$ then an m-dimensional vector space is "larger" than an n-dimensional vector space, which hopefully matches with your intuition.

We end with some useful dimension formulas:

Proposition 21. *Let* $W, U \subset V$ *be subspaces. Then* $\dim(W \oplus U) = \dim(W) + \dim(U)$ *and* $\dim(V/W) = \dim(V) - \dim(W)$.

Proof. Define $T : W \oplus U \to W$ by $x \to x_W$, where the element $x \in W \oplus U$ is written uniquely as $x_W + x_U$ for some $x_W \in W$ and $x_U \in U$. This map is easily see to be linear, and is surjective since $T(w + 0) = w$ for any $w \in W$. We also see that $\ker(T) = U$, so by rank-nullity, we have $\dim(W \oplus U) = \dim(U) + \dim(W)$ as desired.

For the other statement, define $T : V \to V/W$ by $T(v) = v + W$. Then T is linear, and T is clearly surjective by the way it's defined. We also see that $\ker(T) = W$, so that rank-nullity says $\dim(V) = \dim(W) + \dim(V/W)$. □

2.3 The Vector Space $\text{Hom}_F(V, W)$

Definition 2.3.1. A linear transformation $T : V \to W$ is called an **isomorphism** if T is bijective. V and W are called **isomorphic** if there is an isomorphism between them, and we write $V \cong W$.

Definition 2.3.2. Let V and W be F-vector spaces. Then we define $\text{Hom}_F(V, W) = \{T : V \to W : T \text{ is linear}\}$, the set of linear transformations from V to W. If $V = W$, we instead write $\text{End}_F(V)$.

Proposition 22. *$Hom_F(V, W)$ is a subspace of $\mathcal{F}(V, W)$.*

Proof. If $T, U \in \mathrm{Hom}_F(V, W)$ recall that by definition, we have $(T + U)(v) = T(v) + U(v)$ and $(cT)(v) = cT(v)$. To check that $\mathrm{Hom}_F(V, W)$ is a subspace, we need to check that it is non-empty, and that for T, U linear transformations and $c \in F$ that $T + U$ and cT are linear transformations. Note that the 0 function $T(x) = 0$ for all $x \in V$ is certainly linear. If $x, y \in V$ then $(T + U)(x + y) = T(x + y) + U(x + y) = T(x) + U(x) + T(y) + U(y) = (T + U)(x) + (T + U)(y)$. Next, for $c \in F$ we have $(T + U)(cx) = T(cx) + U(cx) = cT(x) + cU(x) = c(T(x) + U(x)) = c(T + U)(x)$, so $T + U$ is a linear transformation which says $T + U \in \mathrm{Hom}_F(V, W)$. For $x, y \in V$, $c, k \in F$, we have $(cT)(x+y) = (cT)(x)+(cT)(y) = cT(x) + cT(y) = (cT)(x) + (cT)(y)$, and $(cT)(kx) = cT(kx) = k(cT(x)) = k(cT)(x)$. This says cT is a linear transformation, so $cT \in \mathrm{Hom}_F(V, W)$ so that $\mathrm{Hom}_F(V, W)$ is a subspace of $\mathcal{F}(V, W)$. □

In most linear algebra books the space $\mathrm{Hom}_F(V, W)$ is denoted as $\mathcal{L}(V, W)$ and $\mathrm{End}_F(V)$ as $\mathcal{L}(V)$, but the above notation is more common elsewhere in mathematics. One of the reasons why finite dimensional vector spaces are so easy to study is that linear transformations between V and W are the same things as functions defined on a basis of V. This reduces much of the study of linear algebra to studying functions defined on a finite set. This is stated precisely in the following form.

Theorem 2.3.3. *Let V be finite dimensional, and let B be a basis of V. Then $Hom_F(V, W) \cong \mathcal{F}(B, W)$. In other words, every linear transformation is determined uniquely by what it does on a basis of V.*

Proof. Suppose that $T : V \to W$ is a linear transformation, and let $B = \{v_1, \ldots, v_n\}$ be a basis of V. For $x \in V$, we may uniquely write $x = c_1v_1 + \ldots + c_nv_n$, so $T(x) = c_1T(v_1) + \ldots + c_nT(v_n)$ by linearity. This defines a function $f : B \to W$ by $f(v_i) = T(v_i)$. Now suppose we have a function $f : B \to W$. Define $T_f : V \to W$ by $T_f(c_1v_1 + \ldots + c_nv_n) = c_1f(v_1) + \ldots + c_nf(v_n)$. We need to check that T_f is a linear transformation, and that it is the only transformation that agrees with f on B. Write $x = c_1v_1 + \ldots + c_nv_n$ and $y = d_1v_1 + \ldots + d_nv_n$. Then $T_f(x + y) = T_f((c_1 + d_1)v_n + \ldots + (c_n + d_n)v_n) = (c_1 + d_1)f(v_1) + \ldots + (c_n + v_n)f(v_n) = c_1f(v_1) + \ldots + c_nf(v_n) + d_1f(v_1) + \ldots + d_nf(v_n) = T_f(x) + T_f(y)$. For $c \in F$, we have $T_f(cx) = T_f((cc_1)v_1 + \ldots + (cc_n)v_n) = (cc_1)f(v_1) + \ldots + (cc_n)f(v_n) = c(c_1f(v_1) + \ldots + c_nf(v_n)) = cT_f(x)$, which shows that T_f is linear. Finally, suppose there is some other linear transformation $T' : V \to W$ such that $T'(v_i) = f(v_i)$. As mentioned above then says for any $x \in V$, $T'(x) = T'(c_1v_1 + \ldots + c_nv_n) = c_1T'(v_1) + \ldots + c_nT'(v_n) = c_1f(v_1) + \ldots c_nf(v_n) = T_f(x)$, i.e. $T' = T_f$ so T_f is the only linear transformation with this property.

Putting this all together, this says the map $G : \mathcal{F}(B, W) \to \mathrm{Hom}_F(V, W)$ with $G(f) = T_f$ is a bijection: it is injective because if $G(f) = G(g)$, then $T_f = T_g$ for all $x \in V$. This then says $T_f(v_i) = f(v_i) = g(v_i) = T_g(v_i)$ for all v_i, so that $f = g$ because they agree on all elements of B. It is surjective because $T \in \mathrm{Hom}_F(V, W)$ defines a map $f : B \to W$ by $f(v_i) = T(v_i)$ and by definition we have $G(f) = T$. It remains to show that G is linear, however this is clear because $G(f + g) = T_{f+g} = T_f + T_g$ because $T_{f+g}(v_i) = (f + g)(v_i) = f(v_i) + g(v_i) = T_f(v_i) + T_g(v_i)$ so T_{f+g} and $T_f + T_g$ agree on B and therefore on all of V. Similarly we see $T_{cf} = cT_f$ for $c \in F$, so G is linear and we are done. □

Theorem 2.3.4. *Two finite dimensional vector spaces are isomorphic if and only if they have the same dimension.*

Proof. Suppose that V, W are finite dimensional with $V \cong W$. Then by definition, there is a bijective linear transformation $T : V \to W$. By rank-nullity, $\text{rank}(T) + \dim(\ker(T)) = \dim(V)$, and since T is a bijection this says $\text{Im}(T) = W$ so $\text{rank}(T) = \dim(W)$ and $\dim(\ker(T)) = 0$, i.e. $\dim(V) = \dim(W)$. Now Suppose that V and W are vector spaces of the same dimension. Let $B = \{v_1, \ldots, v_n\}$ be a basis of V and $B' = \{w_1, \ldots, w_n\}$ be a basis of W. Define $f : B \to W$ by $f(v_i) = w_i$. The previous theorem gives us a linear transformation $T_f : V \to W$. Write $x = c_1v_1 + \ldots + c_nv_n$. Then if $T_f(x) = 0$, this says $c_1T(v_1) + \ldots + c_nT(v_n) = c_1w_1 + \ldots + c_nw_n = 0$, so all $c_i = 0$ because w_i are linearly independent. This says $x = 0$, so $\ker(T) = \{0\}$. Thus, T is injective and therefore bijective, so $V \cong W$. □

2.4 The Matrix of a Linear Transformation

Throughout this section V and W are finite dimensional F-vector spaces of dimensions n and m respectively, with bases $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$.

For any vector $x \in V$, we may uniquely write $x = c_1v_1 + \ldots + c_nv_n$, so the data of the vector x is contained entirely in the list of coefficients of the basis vectors. This gives the following definition.

Definition 2.4.1. The **coordinate representation** of a vector $x = c_1v_1 + \ldots + c_nv_n$ with respect to the basis $\beta = \{v_1, \ldots, v_n\}$ of V is defined by $[x]_\beta = (c_1, \ldots, c_n) \in F^n$.

Example 2.4.2. Let $x = (1, 3) \in \mathbb{R}^2$ and let $\beta = \{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$. Then $x = e_1 + 3e_2$ so $[x]_\beta = (1, 3)$. Set $\gamma = \{(1, 1), (1, -1)\}$. Then $x = 2(1, 1) - (1, -1)$ so $[x]_\gamma = (2, -1)$. With $\alpha = \{(1, 3), (1, 0)\}$ we have $[x]_\alpha = (1, 0)$.

Example 2.4.3. Let $\beta = \{1, x, x^2\}$ be the standard basis of $P_2(\mathbb{R})$. Then with $p(x) = 4 - 3x + 3x^2$, we have $[p(x)]_\beta = (4, -3, 3)$. If $\gamma = \{1, x, \frac{3}{2}x^2 - \frac{1}{2}\}$, we have $[p(x)]_\gamma = (5, -3, 2)$, as $5 - 3x + 2(\frac{3}{2}x^2 - \frac{1}{2}) = p(x)$.

Example 2.4.4. Let $\beta = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ be the standard basis of $M_2(\mathbb{R})$, and let $\gamma = \{E_{11}, E_{12} + E_{21}, E_{22}\}$ be a basis of $\text{Sym}_2(\mathbb{R})$. Let $A = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}$ Then viewed as an element of $M_2(F)$, we may write $[A]_\beta = (2, -1, -1, 5)$, but viewed as an element of $\text{Sym}_2(\mathbb{R})$ we have $[A]_\gamma = (2, -1, 5)$.

Example 2.4.5. View $\mathbb{C}$ as an $\mathbb{R}$-vector space with basis $\beta = \{1, i\}$. Then $x = 3 + 5i$ has $[x]_\beta = (3, 5)$. As a $\mathbb{C}$-vector space, $\mathbb{C}$ has basis $\gamma = \{1\}$, so $[x]_\gamma = 3 + 5i$

Proposition 23. *The map $C_\beta : V \to F^n$ given by $C_\beta(x) = [x]_\beta$ gives an isomorphism $V \cong F^n$.*

Proof. Let $x, y \in V$ with $x = c_1v_1 + \ldots + c_nv_n$ and $y = d_1v_1 + \ldots + d_nv_n$. Then $x + y = (c_1 + d_1)v_1 + \ldots + (c_n + d_n)v_n$. We have $C_\beta(x + y) = (c_1 + d_1, \ldots, c_n + d_n) = (c_1, \ldots, c_n) + (d_1, \ldots, d_n) = C_\beta(x) + C_\beta(y)$. For any $k \in F$, we have $kx = kc_1v_1 + \ldots + kc_nv_n$, so $C_\beta(kx) = (kc_1, \ldots, kc_n) = k(c_1, \ldots, c_n) = kC_\beta(x)$. This proves that C_β is linear. If $C_\beta(x) = 0$, this says that $x = 0v_1 + \ldots + 0v_n = 0$. This says $\ker(C_\beta) = \{0\}$, so C_β is injective and therefore bijective giving $V \cong F^n$. □

Coordinates are one of the best ideas in mathematics, and in linear algebra this is no different. Coordinates give us a way of viewing a vector in an abstract vector space as a more concrete n-tuple of elements of F. In fact, we can do more: using coordinates, we can associate to every linear transformation $T : V \to W$ a matrix $[T]_\beta^\gamma \in M_{m\times n}(F)$. This reduces the study of linear maps from V to W, and therefore linear algebra as a whole, to studying $M_{m\times n}(F)$.

For $x \in V$, write $x = c_1v_1 + \ldots + c_nv_n$, so $T(x) = c_1T(v_1) + \ldots + c_nT(v_n)$. Since the coordinate map $C_\gamma : W \to F^m$ is a linear transformation, this says $[T(x)]_\gamma = c_1[T(v_1)]_\gamma + \ldots + c_n[T(v_n)]_\gamma$. Set $[T(v_i)]_\gamma = (a_{1i}, \ldots, a_{mi})$. Written as a matrix equation, $[T(x)]_\gamma = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$.

Definition 2.4.6. Let $T : V \to W$ be a linear transformation. The **matrix of T** with respect to β and γ is denoted by $[T]_\beta^\gamma$ and is defined by $[T]_\beta^\gamma = \begin{pmatrix} | & & | \\ [T(v_1)]_\gamma & \ldots & [T(v_n)]_\gamma \\ | & & | \end{pmatrix}$. That is, $[T]_\beta^\gamma$ is the matrix whose columns are given by $[T(v_i)]_\gamma$. If $T : V \to V$ and $\beta = \gamma$, then we usually just write $[T]_\beta$.

The definition of the matrix of T says that $[T(x)]_\gamma = [T]_\beta^\gamma[x]_\beta$, and so one can then recover the actual vector $T(x)$ by setting up the corresponding linear combination of basis vectors in γ.

Example 2.4.7. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ with $T(x, y, z) = (x + 3z, -x + 2y + z, x + y + z)$. With $\beta = \{e_1, e_2, e_3\}$, and $\gamma = \{(1, -1, 1), (0, 2, 1), (3, 1, 1)\}$, we see $[T]_\beta = \begin{pmatrix} 1 & 0 & 3 \\ -1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$, and $[T]_\beta^\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Example 2.4.8. Let $T : M_2(F) \to M_2(F)$ be given by $T(A) = A - A^t$. With $\beta = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ the standard basis of $M_2(F)$, we have $[T]_\beta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

Example 2.4.9. Let $\alpha = a + bi \in \mathbb{C}$. View $\mathbb{C}$ as an $\mathbb{R}$-vector space with the standard basis $\beta = \{1, i\}$, and consider the linear transformation $T : \mathbb{C} \to \mathbb{C}$ defined by $T(x) = \alpha x$, the multiplication by α map. Then $[T]_\beta = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. This says any complex number $a + bi$ can be thought of as the matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.

Example 2.4.10. Let $D : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ be the derivative map, and $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{1, x, x^2\}$ be the standard bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$. Then $[D]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$.

Example 2.4.11. Set $V = \mathrm{Span}(\{\sin(x), \cos(x)\}) \subset C^\infty(\mathbb{R})$ and define $T : V \to V$ by $T(f) = 3f + 2f' - f''$. With $\beta = \{\sin(x), \cos(x)\}$, we see that $[T]_\beta = \begin{pmatrix} 4 & 4 \\ 2 & -2 \end{pmatrix}$ because $T(\sin(x)) = 4\sin(x) + 2\cos(x)$ and $T(\cos(x)) = 4\sin(x) - 2\cos(x)$. Using row reduction, one can check the only solution to $[T]_\beta x = 0$ is $x = 0$. This says no non-trivial linear combination of $\sin(x)$ and $\cos(x)$ are solutions the the differential equation $3f + 2f' - f'' = 0$.

Example 2.4.12. Let $T : V \to V$ be linear and suppose W is a T-invariant subspace. Let U be the complement of W in V, so $V = W \oplus U$. If $\{w_1, \dots, w_k\}$ and $\{u_1, \dots, u_\ell\}$ are bases of W and U, then $\{w_1, \dots, w_k, u_1, \dots, u_\ell\}$ is a basis of V. Since $T(W) \subset W$, we may write $T(w_i) = c_{i1}w_1 + \ldots + c_{ik}w_k$, so $[T(w_i)]_\beta = (c_{1i}, \dots, c_{ki}, 0, \dots, 0)$. This gives that $[T]_\beta$ is a block matrix of the form $\begin{pmatrix} A & B \\ O & C \end{pmatrix}$, where A is the $k \times k$ matrix $[c_{ij}]$, O is the $(n-k) \times (n-k)$ zero matrix, and B and C are some matrices of size $(n-k) \times (n-k)$ and $k \times k$ respectively. Therefore having a T-invariant subspace allows one to decompose the matrix of T into an easier to work with block form.

The results of this section that there is a correspondence between matrices and linear transformations can be summed up in the below theorem.

Lemma 2.4.13. *Let $T, U : V \to W$ be linear transformations, and $c \in F$. Then $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$ and $[cT]_\beta^\gamma = c[T]_\beta^\gamma$.*

Proof. By definition, the i-th column of $[T+U]_\beta^\gamma$ is equal to $[(T+U)(v_i)]_\gamma = [T(v_i)+U(v_i)]_\gamma = [T(v_i)]_\gamma + [U(v_i)]_\gamma$ by linearity of the map C_γ. However, this is clearly also the i-th column of the matrix $[T]_\beta^\gamma + [U]_\beta^\gamma$ so $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$. Similarly, the i-th column of the matrix $[cT]_\beta^\gamma$ is given by $[(cT)(v_i)]_\gamma = [cT(v_i)]_\gamma = c[T(v_i)]_\gamma$, which is again the i-th column of the matrix $c[T]_\beta^\gamma$. □

Theorem 2.4.14. *$Hom_F(V, W) \cong M_{m\times n}(F)$. So in particular, $\dim(Hom_F(V, W)) = mn$.*

Proof. Define $F : \mathrm{Hom}_F(V, W) \to M_{m\times n}(F)$ by $F(T) = [T]_\beta^\gamma$. Suppose that $F(T) = F(U)$. Then $[T]_\beta^\gamma = [U]_\beta^\gamma$, so in particular the columns of these matrices are the same so $[T(v_i)]_\gamma = [U(v_i)]_\gamma$ for all i. Translating back to the actual vectors says $T(v_i) = U(v_i)$, i.e. $T = U$ so F is injective. For a matrix $A = \begin{pmatrix} \vert & & \vert \\ x_1 & \dots & x_n \\ \vert & & \vert \end{pmatrix} \in M_{m\times n}(F)$, write $x_i = (a_{1i}, \dots, a_{mi})$. Then define $f : B \to W$ by $f(v_i) = a_{1i}w_1 + \ldots + a_{mi}w_m$. This defines a linear transformation $T : V \to W$ with $T(v_i) = f(v_i)$ so in coordinates, $[T(v_i)]_\gamma = [f(v_i)]_\gamma = x_i$. This then says $[T]_\beta^\gamma = A$, so that F is surjective, so F is a bijection. By the above lemma, F is linear, so F is an isomorphism as desired. The dimension result then follows immediately. □

2.5 Invertibility

Recall that a function $f : X \to Y$ is said to be invertible if there is $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$, and we denote $g = f^{-1}$. From set theory, f is invertible if and only if f is a bijection. For linear transformations $T : V \to W$ and $S : W \to Z$, we denote the composition $S \circ T$ by ST, and clearly then T is an isomorphism if and only if T is invertible. Since linear transformations correspond to matrices, we make a similar definition.

Definition 2.5.1. Let $A \in M_n(F)$. Then A is **invertible** if there is $B \in M_n(F)$ such that $AB = BA = I_n$. If such a matrix exists it's easy to see that it must be unique, and we then write $B = A^{-1}$.

Definition 2.5.2. Let V be a vector space. We define $\mathrm{GL}(V) = \{T \in \mathrm{End}_F(V) : T \text{ is invertible}\}$. Similarly, we set $\mathrm{GL}_n(F) = \{A \in M_n(F) : A \text{ is invertible}\}$.

Proposition 24. *Let $T : V \to W$ and $S : W \to Z$ be linear transformations. Then $ST : V \to Z$ is a linear transformation. If T is invertible, then T^{-1} is a linear transformation.*

Proof. For $x, y \in V$ we have $ST(x+y) = S(T(x+y)) = S(T(x)+T(y)) = ST(x)+ST(y)$, and for $c \in F$, we also have $ST(cx) = S(T(cx)) = S(cT(x)) = cST(x)$ since S, T are linear. Suppose that T is invertible with inverse T^{-1}. For $w, w' \in W$, $T^{-1}(w+w')$ is the vector that maps to $w+w'$ under T. Since T is linear, $T(T^{-1}(w)+T^{-1}(w')) = w+w'$, so $T^{-1}(w+w') = T^{-1}(w)+T^{-1}(w')$. Similarly we see for $c \in F$ that $T^{-1}(cw) = cT^{-1}(w)$ so T^{-1} is linear.

□

Proposition 25. *Let $S : W \to Z$ and $T : V \to W$ be linear transformations, and let α, β, γ be bases of the finite dimensional vector spaces V, W, Z respectively. Then $[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta$*

Proof. Let $\alpha = \{v_1, \ldots, v_n\}$ and $\beta = \{w_1, \ldots, w_k\}$. By definition, the i-th column of $[ST]_\alpha^\gamma$ is given by $[ST(v_i)]_\gamma$. The i-th column of $[S]_\beta^\gamma [T]_\alpha^\beta$ is $S_\beta^\gamma [T(v_i)]_\beta$, so it's sufficient to check these expressions are equal. Write $T(v_i) = c_{1i}w_1 + \ldots + c_{ki}w_k$. Then $ST(v_i) = S(c_{1i}w_1 + \ldots + c_{ki}w_k) = c_{1i}S(w_1) + \ldots + c_{1k}S(w_k)$. Applying C_γ then gives $[ST(v_i)]_\gamma = c_{1i}[S(w_1)]_\gamma + \ldots + c_{1k}[S(w_k)]_\gamma$, which we then recognize as saying $[ST(v_i)]_\gamma = [S]_\beta^\gamma [T(v_i)]_\beta$ as desired.

□

Proposition 26. *Let $T : V \to W$ be a linear transformation, and let β, γ be bases of the finite dimensional vector spaces V and W. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible, and further $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$.*

Proof. If T is invertible, then $T^{-1} : W \to V$ satisfies $TT^{-1} = \mathrm{id}_W$ and $TT^{-1} = \mathrm{id}_V$, so the above says $[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma = I_n$, so that $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$. Conversely, suppose that $[T]_\beta^\gamma$ is invertible. Then the columns of $[T]_\beta^\gamma$ are linearly independent: if not, there are $c_1, \ldots, c_n \in F$ not all 0 such that $c_1[T(v_1)]_\gamma + \ldots + c_n[T(v_n)]_\gamma = 0$, i.e. there is a non-trivial solution to $[T]_\beta^\gamma x = 0$. However, this is impossible because multiplying by the inverse of $[T]_\beta^\gamma$ on the left shows that if the above holds then necessarily $x = 0$. This says that the vectors $[T(v_i)]_\gamma$ in F^n are linearly independent, and as the coordinate mapping is an isomorphism this then implies that the vectors $w_i = T(v_i)$ are linearly independent vectors in W, and therefore are a basis of W. Define a linear transformation $S : W \to V$ by $S(w_i) = v_i$ and extend linearly. By definition, $ST(v_i) = S(w_i) = v_i$, so $ST = \mathrm{id}_V$, and similarly $TS(w_i) = T(v_i) = w_i$, so $TS = \mathrm{id}_W$ so that T is invertible as desired.

□

Knowing if a linear transformation is an isomorphism or not is extremely important – if two vector spaces V and W are isomorphic, this essentially says that W and V are the "same" vector space up to relabeling of the elements, because addition of vectors in one space corresponds uniquely to addition of vectors in the other. This then reduces the study of vector spaces to studying vector spaces up to isomorphism. If T is a linear operator on some vector space V, knowing that T is invertible is very powerful, as will hopefully be demonstrated in the following examples.

Example 2.5.3. Let $V = \mathrm{Span}(\{e^{ax}\sin(bx), e^{ax}\cos(bx)\}) \subset C^\infty(\mathbb{R})$ for $a, b \neq 0$, and let D be the differential operator. Then V is a 2-dimensional D-invariant subspace. With $\beta = \{e^{ax}\sin(bx), e^{ax}\cos(bx)\}$, the matrix $[D|_V]_\beta$ is given by $[D|_V]_\beta = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Since $\det([D|_V]_\beta) = a^2 + b^2 \neq 0$, $D|_V$ is invertible with inverse $A = \begin{pmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ -\frac{b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{pmatrix}$. As $\int f'(x)\,dx = f(x) + C$ and $\frac{d}{dx}\int f(x)\,dx = f(x)$, choosing the choice of constant to be $C = 0$ says the inverse of D is the indefinite integral operator. To integrate say, $\int e^{ax}\sin(bx)\,dx$, we see $A[e^{ax}\sin(bx)]_\beta = (\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2})$, so converting back into vectors of V says $\int e^{ax}\sin(bx)\,dx = \frac{a}{a^2+b^2}e^{ax}\sin(bx) - \frac{b}{a^2+b^2}e^{ax}\cos(bx) + C$ after appending back the arbitrary constant of integration.

Example 2.5.4. ASCII is an encoding standard that associates characters to 7-digit binary strings, which we may think of as elements of $(\mathbb{Z}/2\mathbb{Z})^7$. Fix a matrix $A \in \mathrm{GL}_7(\mathbb{Z}/2\mathbb{Z})$. A simple encryption method is as follows: given a message M, convert each character c of M to ASCII and then convert it into a vector $x_c \in (\mathbb{Z}/2\mathbb{Z})^7$. Encrypt M character-wise by computing Ax_c for all characters, and convert back to text. Since M is invertible, the message can be decrpyted by again converting text to ASCII and multiplying characters by A^{-1}. As an example, the message "TEST" corresponds to the block of binary strings "1010100 1000101 1010011 1010100". With $A = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$, this gets encrypted to the message "gS∼g", which can be decrypted if A or A^{-1} is known.

We showed above that checking the invertibility of a linear operator T is the same as checking the invertibility of its corresponding matrix. We remind the reader of some of many equivalent conditions for checking the latter:

Theorem 2.5.5. *Let $A \in M_n(F)$. Then the following are equivalent:*

(a) A is invertible.

(b) The only solution in F^n to $Ax = 0$ is $x = 0$.

(c) The columns of A are linearly independent.

(d) A is row-equivalent to I_n.

(e) $\det(A) \neq 0$.

(f) The augmented matrix $[A|I]$ is row equivalent to $[I|B]$ for some non-zero matrix B.

2.6 Change of Basis and Similarity

Given a linear operator T on V, the matrix $[T]_\beta$ depends on a choice of basis β of V. Picking a different basis β' will produce a different looking matrix $[T]_{\beta'}$, but it still represents the same operator T. A natural question is given two matrices $A, B \in M_n(F)$, how can one check if they come from the same linear operator in $\mathrm{GL}(V)$?

Pick bases β, γ of V, and consider the identity operator id_V, along with the corresponding matrix $[\mathrm{id}_V]^\gamma_\beta$. This matrix satisfies $[x]_\gamma = [\mathrm{id}_V(x)]_\gamma = [\mathrm{id}_V]^\gamma_\beta [x]_\beta$ for all $x \in V$, or in other words, multiplication by $[\mathrm{id}_V]^\gamma_\beta$ converts the coordinates of the vector x from the basis β to the basis γ.

Definition 2.6.1. Let V be a vector spaces with basis $\beta = \{v_1, \dots, v_n\}$, and let $\gamma = \{v'_1, \dots, v'_n\}$ be another basis. The **change of basis matrix** from β to γ, denoted S^γ_β, is the matrix $[\mathrm{id}_V]^\gamma_\beta$. Explicitly, $S^\gamma_\beta = \begin{pmatrix} | & & | \\ [v_1]_\gamma & \dots & [v_n]_\gamma \\ | & & | \end{pmatrix}$.

Since id_V is invertible, this says S^γ_β is invertible, and has inverse matrix $[\mathrm{id}_V]^\beta_\gamma = S^\beta_\gamma$.

Theorem 2.6.2. *Let β, γ be two bases of a finite dimensional vector space V, and let $T \in GL(V)$. Then $[T]_\gamma = S^\gamma_\beta [T]_\beta S^\beta_\gamma$, and $[T]_\gamma S^\gamma_\beta = S^\gamma_\beta [T]_\beta$. In otherwords, the following diagram*

$$\begin{tikzcd}
{[x]_\beta} \arrow[r, "{[T]_\beta}"] \arrow[d, "S^\gamma_\beta"] & {[T(x)]_\beta} \arrow[d, "S^\gamma_\beta"] \\
{[x]_\gamma} \arrow[r, "{[T]_\gamma}"] & {[T(x)]_\gamma}
\end{tikzcd}$$

commutes.

Proof. Since composition of linear transformations corresponds to multiplication by their corresponding matrices, we see $[T]_\gamma = [\mathrm{id}_V \circ T \circ \mathrm{id}_V]^\gamma_\gamma = [\mathrm{id}_V]^\gamma_\beta [T]_\beta [\mathrm{id}_V]^\beta_\gamma = S^\gamma_\beta [T]_\beta S^\beta_\gamma$. Since S^γ_β is invertible with inverse S^β_γ, multiplication on the left gives $S^\beta_\gamma [T]_\gamma = [T]_\beta S^\beta_\gamma$ as desired. □

Example 2.6.3. Let $\beta = \{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$ and $\gamma = \{(1,1),(1,2)\}$ be another basis. The change of basis matrix S^β_γ is $S^\beta_\gamma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. To compute S^γ_β, we take the inverse to find $S^\gamma_\beta = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. To compute $[e_1]_\gamma$, we see $[e_1]_\gamma = S^\gamma_\beta [e_1]_\beta = S^\gamma_\beta e_1 = (2,-1)$, so that $(1,0) = 2(1,1) - (1,2)$.

Example 2.6.4. Let $\beta = \{1, x, x^2\}$ and $\gamma = \{1, x, \frac{3}{2}x^2 - \frac{1}{2}\}$ be bases of $P_2(\mathbb{R})$. Then $S^\beta_\gamma = \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{pmatrix}$, and one can compute $S^\gamma_\beta = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}$. Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be defined by $T(f)(x) = xf'(x)$. Then $[T]_\beta = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ and the change of basis formula says $[T]_\gamma = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

Example 2.6.5. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x,y,z) = (2z, -2x+3y+2z, -x+3z)$. Let $\beta = \{e_1, e_2, e_3\}$ be the standard basis of $\mathbb{R}^3$, and let $\gamma = \{(2,1,1),(1,0,1),(0,1,0)\}$ be another basis. Then $[T]_\beta = \begin{pmatrix} 0 & 0 & 2 \\ -2 & 3 & 2 \\ -1 & 0 & 3 \end{pmatrix}$. We see $S^\beta_\gamma = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, and $S^\gamma_\beta = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 0 & 2 \\ -1 & 1 & 1 \end{pmatrix}$, so the change of basis formula says $[T]_\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. With respect to the new basis γ, this says that T acts along each γ-direction by scaling. Having a basis where an operator is diagonal is extremely useful, as it allows one to easily compute values of compositions. For example, to compute $T^n(1,2,3)$, we compute $[T^n(1,2,3)]_\gamma = [T^n]_\gamma [(1,2,3)]_\gamma = [T]^n_\gamma [(1,2,3)]_\gamma$. We have $[(1,2,3)]_\gamma = S^\gamma_\beta [(1,2,3)]_\beta = (-2,5,4)$, so $[T^n(1,2,3)]_\gamma = [T]^n_\gamma (-2,5,4)^T = (-2, 5 \cdot 2^n, 4 \cdot 3^n)$. This says $T^n(1,2,3) = -2(2,1,1) + 5 \cdot 2^n (1,0,1) + 4 \cdot 3^n (0,1,0) = (-4 + 5 \cdot 2^n, -2 + 4 \cdot 3^n, -2 + 5 \cdot 2^n)$.

Example 2.6.6. Consider $P_3(\mathbb{R})$, and set $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{\binom{x}{0}, \dots, \binom{x}{3}\}$, where $\binom{x}{0} = 1$ and for $k \geq 1$ we have $\binom{x}{k} = \frac{x(x-1)\dots(x-k+1)}{k!}$. Then γ is a basis, because each polynomial in γ has a different degree. The change of basis matrix S^β_γ is given by $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{6} \end{pmatrix}$, and one can check that $S_\beta^\gamma = \begin{pmatrix}1&0&0&0\\0&1&1&1\\0&0&2&6\\0&0&0&6\end{pmatrix}$. Then $[x^3]_\gamma = S_\beta^\gamma[x^3]_\beta = (0,1,6,6)$, so $x^3 = \binom{x}{1}+6\binom{x}{2}+6\binom{x}{3}$. As an application, $\sum_{k=1}^{n-1} k^3 = \sum_{k=1}^{n-1}\binom{k}{1}+6\binom{k}{2}+6\binom{k}{3} = \binom{n}{2}+6\binom{n}{3}+6\binom{n}{4} = (\frac{n(n-1)}{2})^2$, which follows from the identity $\sum_{k=1}^{n-1}\binom{k}{r} = \binom{n}{r+1}$, easily proven by induction.

The theorem from above leads us to the following definition:

Definition 2.6.7. For $A, B \in M_n(F)$, we say that A and B are **similar** and write $A \sim B$ if there exists $P \in \mathrm{GL}_n(F)$ such that $A = PBP^{-1}$.

Since $S_\gamma^\beta = (S_\beta^\gamma)^{-1}$, this says that for any choice of bases β, γ of V the matrices $[T]_\beta$ and $[T]_\gamma$ are similar. The following observation is easy to verify:

Proposition 27. *Similarly is an equivalence relation on $M_n(F)$.*

We showed that changing the basis of V from β to γ yields similar matrices $[T]_\beta$ and $[T]_\gamma$. The converse is true as well:

Theorem 2.6.8. *$A \sim B$ in $M_n(F)$ if and only if there are bases β, γ of V and a linear transformation T such that $A = [T]_\beta$ and $B = [T]_\gamma$. That is, similar matrices correspond to the same linear transformation under potentially different bases.*

Proof. Suppose that $A \sim B$ in $M_n(F)$, so there is $P \in \mathrm{GL}_n(F)$ such that $A = PBP^{-1}$, and let $\beta = \{v_1, \ldots, v_n\}$ be any basis of V. We have seen that we may choose T such that $[T]_\beta = A$, so it remains to find γ such that $[T]_\gamma = B$. To do this, we would like to think of P as some change of basis matrix. Define w_i such that $[w_i]_\beta = Pe_i$, where e_i are the standard basis vectors of F^n. As P is invertible, its columns are linearly independent, and because the coordinate map C_β is an isomorphism, w_i are also linearly independent so that $\gamma = \{w_1, \ldots, w_n\}$ is a basis of V. Then S_γ^β is the matrix with columns $[w_i]_\beta = Pe_i$, so $S_\gamma^\beta = P$, and $S_\beta^\gamma = P^{-1}$. Since $A = PBP^{-1}$, this says $B = P^{-1}AP = S_\beta^\gamma[T]_\beta S_\gamma^\beta = [T]_\gamma$ as desired. The backwards direction was proven above.

□

The conjugacy classes of matrices in $M_n(F)$ under similarly correspond to the *distinct* linear operators $T \in \mathrm{GL}(V)$, regardless of choice of basis. Therefore if one cares only about the different types of operators that arise on V, the importance of studying matrices up to similarity is self evident. To be able to distinguish between conjugacy classes, it's helpful to known some quantities that are invariant under similarity.

Lemma 2.6.9. *Let $A, B \in M_n(F)$. Then $tr(AB) = tr(BA)$.*

Proof. Write $A = (a_{ij})$ and $B = (b_{ij})$. Then $\mathrm{tr}(AB) = \sum_{i=1}^n (AB)_{ii} = \sum_{i=1}^n \sum_{k=1}^n a_{ik}b_{ki}$. On the other hand, $\mathrm{tr}(BA) = \sum_{i=1}^n (BA)_{ii} = \sum_{i=1}^n \sum_{k=1}^n b_{ik}a_{ki}$. By renaming variables i and k, we have $\sum_{i=1}^n \sum_{k=1}^n b_{ik}a_{ki} = \sum_{k=1}^n \sum_{i=1}^n b_{ki}a_{ik}$, and by swapping the order of summation this equals $\sum_{i=1}^n \sum_{k=1}^n a_{ik}b_{ki}$ as desired. □

Proposition 28. *Let $A, B \in M_n(F)$, with $A \sim B$. Then $tr(A) = tr(B)$ and $det(A) = det(B)$.*

Proof. As $A \sim B$, there is $P \in M_n(F)$ such that $A = PBP^{-1}$. By the lemma, $\operatorname{tr}(A) = \operatorname{tr}(P(BP^{-1})) = \operatorname{tr}((BP^{-1})P) = \operatorname{tr}(B)$. Similarly, we find that $\det(A) = \det(PBP^{-1}) = \det(P)\det(B)\det(P^{-1}) = \det(B)\det(P)\det(P^{-1}) = \det(B)\det(I_n) = \det(B)$ by properties of the determinant. □

Since similarity corresponds to change of basis, this allows us to define these quantities for a linear operator.

Definition 2.6.10. Let V be finite dimensional. For $T \in \mathrm{GL}(V)$ we define the **trace** of T and the **determinant** of T to be the quantities $\operatorname{tr}([T]_\beta)$ and $\det([T]_\beta)$ for any choice of basis β of V.

Example 2.6.11. The matrices $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 2 \\ 1 & -3 \end{pmatrix}$ are not similar, because $\operatorname{tr}(A) = 5$ while $\operatorname{tr}(B) = -3$. However, both $\det(A) = \det(B) = -2$.

Recall that the rank of a linear transformation T was defined as the dimension of it's image. We can also similarly define rank in terms of a matrix representation of T:

Proposition 29. *Let $A, B \in M_n(F)$ and suppose that $A \sim B$. Then $rank(A) = rank(B)$.*

Proof. Write $A = PBP^{-1}$ for some P. Define $T : \operatorname{Im}(A) \to F^n$ by $T(x) = P^{-1}x$. Then T is injective because P is invertible, so $\dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(T))$. We also see that if $y \in \operatorname{Im}(T)$, then $y = P^{-1}x$ for some $x \in \operatorname{Im}(A)$. Write $x = Az$, so $y = P^{-1}(Az) = B(P^{-1}z) \in \operatorname{Im}(B)$. This gives $\operatorname{rank}(A) \leq \operatorname{rank}(B)$. Similarly with $S : \operatorname{Im}(B) \to F^n$ defined by $S(x) = Px$, we find $\operatorname{rank}(B) \leq \operatorname{rank}(A)$, so that $\operatorname{rank}(A) = \operatorname{rank}(B)$. □

Chapter 3

Diagonalization

After developing the basic theory of linear transformations, we turn our attention to the study of linear operators. As we saw, a change of basis of V from β to β' corresponds to conjugation of the matrix $[T]_\beta$ by some invertible change of basis matrix $S_\beta^{\beta'}$. A natural question is if there is a "best" basis β to pick so that $[T]_\beta$ will be as easy to understand as possible. The best we could hope for in general is that $[T]_\beta$ is a diagonal matrix, so the question becomes if it is possible to find a basis β of V such that $[T]_\beta$ is diagonal. Answering this question will be the primary purpose of this handout.

Throughout this document, V will denote a vector space over an arbitrary field F and $T : V \to V$ will denote a linear operator.

3.1 Basic Definitions

Definition 3.1.1. An **eigenvector** of T is a non-zero vector $v \in V$ such that $T(v) = \lambda v$ for some $\lambda \in F$. The number λ is called an **eigenvalue** of T. We sometimes refer to the data (v, λ) as an **eigenpair**.

Our first order of business is to determine what the possible eigenvalues of a linear operator are. When V is finite dimensional, this is quite easily done using the theory of determinants. We remind the reader of the following definition:

Definition 3.1.2. Let V be an n dimensional vector space. The **determinant** of a linear operator $T : V \to V$ denoted $\det(T)$ is defined as $\det([T]_\beta)$ for any basis β of V.

Elementary properties of the determinant show that similar matrices have the same determinant, so the above definition actually is independent of a choice of basis and so the notation makes sense.

Theorem 3.1.3. *Let V be an n dimensional vector space. Then $\lambda \in F$ is an eigenvalue of T if and only if* $\det(\lambda \cdot I_V - T) = 0$. *In terms of matrices, λ is an eigenvalue of T if and only if* $\det(\lambda \cdot I_n - [T]_\beta) = 0$ *for any basis β of V.*

Proof. Pick a basis β of V. Suppose that $\lambda \in F$ is an eigenvalue of T with eigenvector v. Then $(\lambda \cdot I_V - T)(v) = 0$, i.e. the operator $\lambda \cdot I_V - T$ is not invertible, and from the theory of matrices, this says $[\lambda \cdot I_V - T]_\beta$ is not invertible. Therefore $\det([\lambda \cdot I_V - T]_\beta) = 0$, so by definition this says that $\det(\lambda \cdot I_V - T) = 0$. Conversely, if $\det(\lambda \cdot I_V - T) = 0$, then $\lambda \cdot I_V - T$ is not invertible, so there is some vector v in the kernel of $\lambda \cdot I_V - T$, i.e. a vector v such that $T(v) = \lambda v$ so that λ is an eigenvalue of T. □

The above says that checking if λ is an eigenvalue of T is equivalent to finding a root of the polynomial $\det(x \cdot I_V - T)$. We give this polynomial a name:

Definition 3.1.4. The polynomial $p_T(x) = \det(x \cdot I_V - T)$ is called the **characteristic polynomial** of T.

Restated in the new definition, we have the following:

Theorem 3.1.5. *Let V be an n dimensional vector space. $\lambda \in F$ is an eigenvalue of $T : V \to V$ if and only if λ is a root of the characteristic polynomial $p_T(x)$.*

Notice that since the determinant of a linear operator does not depend on a choice of basis, the characteristic polynomial is independent of such a choice as well, and so it is well defined.

Before moving on, we make a few remarks: notice that the definition of an eigenvalue depends on which field we view V as a vector space over. If F is algebraically closed, then every linear operator $T : V \to V$ has an eigenvalue, because then the characteristic polynomial of T necessarily has a root in F. If F is not algebraically closed, it may be possible for an operator to not have an eigenvalue. Additionally, the definition of an eigenvalue makes sense for infinite dimensional vector spaces, even if our criterion for easily finding eigenvalues only works for finite dimensional vector spaces. Some of the examples below will illustrate this.

Example 3.1.6. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (y, -5x+4y+z, -x+y+z)$. Then with β the standard basis, we see $[T]_\beta = \begin{pmatrix} 0 & 1 & 0 \\ -5 & 4 & 1 \\ -1 & 1 & 1 \end{pmatrix}$. Then $p_T(x) = (x-1)(x-2)^2$, so T has eigenvalues 1 and 2. One can check T has eigenvectors $v_1 = (1, 1, 2)$ and $v_2 = (1, 2, 1)$ corresponding to the eigenvalues 1 and 2 respectively.

Example 3.1.7. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation by some angle $\theta \in (0, 2\pi)$. Then T has no eigenvectors, because no vector $v \in \mathbb{R}^2$ is scaled along the same direction by T. Explicitly, with $\beta = \{e_1, e_2\}$ the standard basis of $\mathbb{R}^2$, one can check that $[T]_\beta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. The characteristic polynomial of T is given by $p_T(x) = x^2 - 2\cos(\theta)x + 1$, and the quadratic formula says this has no real roots.

Example 3.1.8. Write $V = W \oplus W'$ for some subspaces W, W' of V. Let $P = \pi_W$ be the projection onto W, so that $P^2 = P$. If (λ, v) is an eigenpair for P, then $\lambda^2 v = P^2 v = Pv = \lambda v$, so that $\lambda^2 = \lambda$ says that $\lambda = 0, 1$ are the only possible eigenvalues of P. For any $w \in W$, $P(w) = w$, and for $w' \in W'$, $P(w') = 0$, so w, w' are eigenvectors corresponding to $0, 1$ respectively.

Example 3.1.9. Let $T : C^{\infty}(\mathbb{R}) \to C^{\infty}(\mathbb{R})$ be the derivative map, $T(f) = f'$. An eigenvector of T is a function f such that $f' = \lambda f$ for some $\lambda \in \mathbb{R}$. From calculus, we know that such functions are of the form $ce^{\lambda t}$ for some $c \in \mathbb{R}$. This says the exponential functions $ce^{\lambda t}$ are eigenvectors of the derivative operator with eigenvalues $\lambda \in \mathbb{R}$. This is of fundamental importance in the theory of linear differential equations.

Example 3.1.10. Let $L : F^{\infty} \to F^{\infty}$ be the left shift operator, i.e. $L((a_1, a_2, \ldots)) = (a_2, a_3, \ldots)$. An eigenvector of L is a sequence $(a_1, a_2, \ldots)$ such that $(a_2, a_3, \ldots) = \lambda(a_1, a_2, \ldots)$ for some $\lambda \in F$. This says $a_2 = \lambda a_1$, $a_3 = \lambda a_2 = \lambda^2 a_1$, and by induction, that $a_n = \lambda^{n-1} a_1$. Let $\sigma = \{a_n\}$ be a geometric sequence, that is a sequence defined by $a_n = cr^{n-1}$ for some $c, r \in F$. Then we see that an eigenvector of L is a geometric sequence, and any such geometric sequence σ is an eigenvector with eigenvalue r.

3.2 Properties of the Characteristic Polynomial

Before moving on, it will be useful to know some basic properties of the characteristic polynomial $p_T(x)$ of a linear operator T.

Proposition 30. *Let V be an n dimensional vector space and $T : V \to V$ a linear operator. Then the characteristic polynomial $p_T(x)$ is a monic degree n polynomial.*

Proof. Pick a basis β of V, so by definition $p_T(x) = \det(x \cdot I_n - [T]_\beta)$, and write $[T]_\beta = [a_{ij}]$. Then $x \cdot I_n - [T]_\beta$ has entries $x - a_{ii}$ along the diagonal and $-a_{ij}$ elsewhere. Expanding the determinant out using co-factors along the first row, we see that we can write $p_T(x) = (x - a_{11}) \cdots (x - a_{nn}) + q(x)$ where $q(x)$ is a polynomial of degree at most $n - 2$. It's then clear that $p_T(x)$ is a monic degree n polynomial. □

Proposition 31. *Let V be an n dimensional vector space and T a linear operator with characteristic polynomial $p_T(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$. Then $a_{n-1} = -\operatorname{tr}(T)$ and $a_0 = (-1)^n \det(T)$.*

Proof. Write $p_T(x) = (x - a_{11}) \cdots (x - a_{nn}) + q(x)$ where $q(x)$ has degree at most $n - 2$ as above. Then the coefficient of x^{n-1} comes entirely from the coefficient of x^{n-1} in the product $(x - a_{11}) \ldots (x - a_{nn})$, which is equal to the negative sum of its roots, i.e. $-(a_{11} + \ldots + a_{nn}) = -\operatorname{tr}(T)$. We also see $a_0 = p_T(0) = \det(0 \cdot I_n - T) = \det(-T) = (-1)^n \det(T)$. □

Example 3.2.1. A particularly useful special case is when V is 2 dimensional, so for a linear operator T we have $p_T(x) = x^2 - \operatorname{tr}(T)x + \det(T)$.

The above result says that the characteristic polynomial $p_T(x)$ record both the trace and determinant of T. Since the roots of $p_T(x)$ are eigenvalues of T, this gives the following important relations:

Corollary 3.2.2. *Let V be an n dimensional vector space and T a linear operator. Suppose that the eigenvalues of T in some algebraic closure $\overline{F}$ of F are $\lambda_1, \ldots, \lambda_n$ (counted with multiplicity). Then* $\operatorname{tr}(T) = \lambda_1 + \ldots + \lambda_n$ *and* $\det(T) = \lambda_1 \cdots \lambda_n$.

Proof. In $\overline{F}[x]$, $p_T(x)$ factors as $(x - \lambda_1) \cdots (x - \lambda_n)$. Expanding out the product then says the coefficient of x^{n-1} is $-(\lambda_1 + \ldots + \lambda_n)$ and the constant term is $(-1)^n(\lambda_1 \cdots \lambda_n)$. □

3.3 Diagonalization

We now use the theory of eigenvalues to answer our main question.

Definition 3.3.1. A linear operator T is called *diagonalizable* if there exists a basis β of V such that $[T]_\beta$ is a diagonal matrix.

Necessarily, we see that if such a basis $\beta = \{v_1, \ldots, v_n\}$ exists, if $[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$ that $T(v_i) = \lambda_i v_i$ for all i, so that β is a basis of eigenvectors. This can be rephrased using the language of the previous section as such:

Theorem 3.3.2. *A linear operator T is diagonalizable if and only if there is a basis β of V consisting of eigenvectors of T.*

In order to determine when such a basis exists, we will utilize the following key result:

Proposition 32. *Suppose $v_1, \ldots, v_n$ are eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $\{v_1, \ldots, v_n\}$ is linearly independent.*

Proof. We prove this by induction. If $n = 1$ the statement follows immediately since $\{v_1\}$ is linearly independent. Assume that the statement is true for any collection of $n - 1$ eigenvectors that correspond to distinct eigenvalues. Suppose that $c_1 v_1 + \ldots + c_n v_n = 0$, so that applying T says $c_1 \lambda_1 v_1 + \ldots + c_n \lambda_n v_n = 0$. Multiply the first equation by λ_n and subtract to see $c_1(\lambda_1 - \lambda_n) v_1 + \ldots + c_{n-1}(\lambda_{n-1} - \lambda_n) v_{n-1} = 0$. By induction hypothesis, the vectors $\{v_1, \ldots, v_{n-1}\}$ are linearly independent, and since all eigenvalues are distinct this forces $c_i = 0$ for $1 \leq i \leq n - 1$. This then immediately gives $c_n = 0$, so that $\{v_1, \ldots, v_n\}$ is linearly independent. By induction, we are done. □

Corollary 3.3.3. *Let V be an n dimensional vector space and T a linear operator. If T has n distinct eigenvalues, then T is diagonalizable. If $p_T(x)$ factors into distinct linear factors in $F[x]$, then T is diagonalizable.*

Proof. If T has n distinct eigenvalues, then the associated eigenvectors are a set of n linearly independent vectors in V, hence a basis. Saying that $p_T(x)$ splits into distinct linear factors is the same as saying that T has distinct eigenvalues. □

Example 3.3.4. The converse to the above statement is not necessarily true. For example, the identity operator I_V is diagonalizable, but has characteristic polynomial $(x - 1)^n$.

Proposition 33. *Let V be an n dimensional vector space and T a linear operator. Then if T is diagonalizable, $p_T(x)$ factors into a product of (not necssesarily distinct) linear factors in $F[x]$.*

Proof. Suppose that T is diagonalizable. Let β be a basis of V consisting of eigenvalues $\lambda_1, \ldots, \lambda_n$ of T. Then $[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$, and $\det(xI_n - [T]_\beta) = (x-\lambda_1)\cdots(x-\lambda_n)$ is a product of linear factors in $F[x]$. □

Example 3.3.5. The converse of the above statement is not necessarily true. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x,y) = (y,0)$. Then it's easy to see $p_T(x) = x^2$, so the only eigenvalue of T is 0. However, T is not diagonalizable: rank-nullity says $\dim(\ker(T)) = 1$, so that there is no possible basis of eigenvectors for T.

The above examples show that the characteristic polynomial is not strong enough to detect when an operator is diagonalizable or not, and we saw that what broke was that nothing can be said when the characteristic polynomial has repeated roots. This leads us to the following definitions, which will end up giving a test for diagonalizability.

Definition 3.3.6. Let λ be an eigenvalue of T. The **eigenspace of** λ, denoted E_λ, is defined as $E_\lambda = \ker(T - \lambda \cdot I_V)$. The **algebraic multiplicity** of λ is its multiplicity as a root of $p_T(x)$. The **geometric multiplicity** of λ is $\dim(E_\lambda)$.

Proposition 34. *The geometric multiplicity of an eigenvalue λ of T is at most its algebraic multiplicity.*

Proof. Suppose that $\dim(E_\lambda) = k$. Pick a basis $\{v_1, \ldots, v_k\}$ of E_λ and extend to a basis $\beta = \{v_1, \ldots, v_k, w_1, \ldots, w_m\}$ of V. Then $[T]_\beta$ is a block matrix given by $[T]_\beta = \begin{pmatrix} \lambda \cdot I_k & A \\ 0 & B \end{pmatrix}$ where 0 is the $m \times k$ zero matrix, and A, B have dimensions $k \times m$ and $m \times m$ respectively. Then $x \cdot I_n - [T]_\beta = \begin{pmatrix} (x-\lambda)\cdot I_k & -A \\ 0 & x \cdot I_m - B \end{pmatrix}$, so $p_T(x) = \det(x \cdot I_n - [T]_\beta) = \det((x-\lambda) \cdot I_k)\det(x \cdot I_m - B) = (x-\lambda)^k g(x)$ where $g(x) = \det(x \cdot I_m - B)$. This says $(x-\lambda)^k$ divides $p_T(x)$, so the algebraic multiplicity of λ is at least k as desired. □

Theorem 3.3.7. *Let V be an n dimensional vector space. Let T have distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ with algebraic multiplicities $e_1, \ldots, e_k$, so $p_T(x) = (x-\lambda_1)^{e_1}\cdots(x-\lambda_k)^{e_k}$, and $e_1 + \ldots + e_k = n$. Then T is diagonalizable if and only if $V = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k}$.*

Proof. If T is diagonalizable, then V has a basis of eigenvectors of T, so that $V = E_{\lambda_1} + \ldots + E_{\lambda_k}$. Since eigenvectors from different eigenvalues are linearly independent, if $x_1 + \ldots + x_k = 0$ with $x_i \in E_{\lambda_i}$, writing x_i in terms of basis vectors of E_{λ_i} shows all $x_i = 0$, so that the sum is direct. Conversely, if $V = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k}$, then the union of bases for E_{λ_i} is a basis of V. Since each vector in E_{λ_i} is an eigenvector of T, this says V has a basis of eigenvectors for T, i.e. T is diagonalizable. □

Corollary 3.3.8. *Let T be as above. Then T is diagonalizable if and only if for each eigenvalue λ_i of T the algebraic multiplicity and geometric multiplicity of λ_i are equal.*

Proof. If the algebraic and geometric multiplicity of λ_i are equal for all i, this says $\dim(E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k}) = e_1 + \ldots + e_k = n$, so $E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k} = V$ says T is diagonalizable. Conversely if $\dim(E_{\lambda_i}) < e_i$ for some i, then $\dim(E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k}) < n$, so $E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k} \neq V$ says T is not diagonalizable. □

Example 3.3.9. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (y, -5x + 4y + z, -x + y + z)$ be the example from before. Then we saw $p_T(x) = (x-1)(x-2)^2$, so 1 has algebraic multiplicity 1 and 2 has algebraic multiplicity 2. To check if T is diagonalizable, we need to see if the eigenspace E_2 is 2-dimensional. We see that $T - 2 \cdot I_V$ has matrix representation with respect to the standard basis given by $\begin{pmatrix} -2 & 1 & 0 \\ -5 & 2 & 1 \\ -1 & 1 & -1 \end{pmatrix}$. Notice that $-2(1, 2, 1) + (0, 1, -1) = (-2, -5, -1)$, so that the first column is a linear combination of the other two. The second and third columns are obviously linearly independent, so that $\operatorname{rank}(T - 2 \cdot I_V) = 2$ says $\dim(\ker(T - 2 \cdot I_V)) = 1$, so that T is not diagonalizable.

Example 3.3.10. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (-9x + 4y + 4z, -8x + 3y + 4z, -16x + 8y + 7z)$. With $\beta = \{e_1, e_2, e_3\}$ the standard basis of $\mathbb{R}^3$, we have $[T]_\beta = \begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}$. One can compute that $p_T(x) = (x - 3)(x + 1)^2$, so that T has eigenvalues 3 and -1 with algebraic multiplicities 1 and 2 respectively. The operator $T + I_V$ has matrix representation $\begin{pmatrix} -8 & 4 & 4 \\ -8 & 4 & 4 \\ -16 & 8 & 8 \end{pmatrix}$, so $\dim(E_{-1}) = 2$, so that T is diagonalizable. Using row reduction, a basis of E_{-1} is given by $v_1 = (1, 0, 2)$ and $v_2 = (1, 2, 0)$. The operator $T - 3 \cdot I_V$ has matrix representation $\begin{pmatrix} -12 & 4 & 4 \\ -8 & 0 & 4 \\ -16 & 8 & 4 \end{pmatrix}$, and a basis of E_3 is given by $v_3 = (1, 1, 2)$. Then $\beta' = \{v_1, v_2, v_3\}$ is a basis of $\mathbb{R}^3$ consisting of eigenvectors for T. The change of basis matrix $S_{\beta' \to \beta}$ is the matrix whose columns are these eigenvectors, i.e. $S_{\beta' \to \beta} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 2 & 0 & 2 \end{pmatrix}$, with inverse $S_{\beta \to \beta'} = \begin{pmatrix} 2 & -1 & -1/2 \\ 1 & 0 & -1/2 \\ -2 & 1 & 1 \end{pmatrix}$. The change of basis formula says $[T]_\beta = S_{\beta' \to \beta}[T]_{\beta'} S_{\beta \to \beta'}$, and we see that $[T]_{\beta'} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, which gives a factorization of $[T]_\beta$.

Chapter 4

Inner Product Spaces

When we started our study of vector spaces, we had a goal in mind: find objects that generalized the algebraic structure on Euclidean space $\mathbb{R}^n$. However, if the ultimate goal of linear algebra is to fully generalize Euclidean space, there's something major that still hasn't been abstracted: the *geometry* of $\mathbb{R}^n$. The definition of an abstract vector space V does not include notions of length, distance, or angles, and therefore no concept of geometry. In order for a vector space to truly "act" Euclidean, we need to add more structure.

4.1 Basic Definitions

Throughout this document, we assume $F = \mathbb{R}$ or $F = \mathbb{C}$, and V is an inner product space over F.

Definition 4.1.1. An **inner product** $\langle -, - \rangle : V \times V \to F$ is a function that satisfies the following properties:

1. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$ for all $x, y, z \in V$
2. $\langle cx, y \rangle = c\langle x, y \rangle$ for all $x, y \in V$ and $c \in F$
3. $\overline{\langle x, y \rangle} = \langle y, x \rangle$ for all $x, y \in V$
4. $\langle x, x \rangle > 0$ for all $x \neq 0 \in V$.

An **inner product space** is a pair $(V, \langle -, - \rangle)$, i.e. a vector space V with a choice of inner product on V. From the conjugate symmetry of the inner product, we deduce the following basic properties:

Proposition 35. *Let V be an inner product space. Then the following hold:*

(a) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ *for all* $x, y, z \in V$.

(b) $\langle x, cy \rangle = \overline{c}\langle x, y \rangle$ *For all* $x, y \in V$ *and* $c \in F$.

(c) $\langle x, 0 \rangle = \langle 0, x \rangle = 0$ *for all* $x \in V$.

(d) $\langle x, x\rangle = 0$ *if and only if* $x = 0$.

(e) If $\langle x, y\rangle = \langle x, z\rangle$ *for all* $x \in V$*, then* $y = z$.

Proof. These are routine verifications and are omitted. □

The idea of an inner product is to generalize the dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$. Having an inner product gives us a notion of length:

Definition 4.1.2. The **norm** of $x \in V$, denoted $\|x\|$, is defined by $\|x\| = \sqrt{\langle x, x\rangle}$.

Proposition 36. *Let* V *be an inner product space. Then the norm* $\|\cdot\|$ *satisfies the following properties:*

(a) $\|cx\| = |c|\|x\|$ *for all* $x \in V$*,* $c \in F$.

(b) $\|x\| = 0$ *if and only if* $x = 0$.

(c) (Cauchy-Schwarz) For all $x, y \in V$ $|\langle x, y\rangle| \leq \|x\|\|y\|$ *and equality holds if and only if* $x = cy$ *for some* $c \in F$.

(d) (Triangle inequality) For all $x, y \in V$*,* $\|x + y\| \leq \|x\| + \|y\|$ *and equality holds if and only if* $x = cy$ *for some* $c \in F$.

Proof.

(a) $\|cx\| = \langle cx, cx\rangle = c\bar{c}\langle x, x\rangle = |c|\|x\|$.

(b) This is (d) from the above proposition.

(c) If $y = 0$ this is obvious, so assume $y \neq 0$. For $c \in F$, we have $0 \leq \|x - cy\|^2 = \langle x - cy, x - cy\rangle = \langle x, x\rangle - \bar{c}\langle x, y\rangle - c\langle y, x\rangle + c\bar{c}\langle y, y\rangle$ by expanding out the inner product. In particular, setting $c = \frac{\langle x,y\rangle}{\langle y,y\rangle}$, this becomes $0 \leq \|x\|^2 - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^2}$, so that $0 \leq \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}$. This then gives $|\langle x, y\rangle| \leq \|x\|\|y\|$ as desired. The case of equality is left as an exercise.

(d) We have $\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle = \|x\|^2 + 2\text{Re}(\langle x, y\rangle) + \|y\|^2$ by expanding out the inner product and using conjugate symmetry. Since $2\text{Re}(\langle x, y\rangle) \leq 2|\langle x, y\rangle| \leq 2\|x\|\|y\|$ by Cauchy-Schwarz, we then see $\|x + y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$, so take a square root to finish up. The equality case is again left as an exercise.

□

We now give some standard examples of inner product spaces.

Example 4.1.3. If V has an inner product $\langle -, -\rangle$ then for any subspace W of V, $\langle -, -\rangle$ is still an inner product on W.

Example 4.1.4. Set $V = \mathbb{R}^n$ and let $\cdot$ be the usual dot product, $(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1b_1 + \ldots + a_nb_n$. This makes V a real inner product space. If instead $V = \mathbb{C}^n$, we define the dot product to be $(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1\overline{b_1} + \ldots + a_n\overline{b_n}$. This makes V a complex inner product space. As an example in $\mathbb{C}^2$, we have $(1+2i, 3-i) \cdot (2, i) = 2(1+2i) - i(3-i) = 1 - i$.

Example 4.1.5. If V is a finite dimensional F-vector space, we can always give V the structure of an inner product space as follows. Say that $\dim_F(V) = n$, and fix an isomorphism $\varphi : V \to F^n$. Define an inner product $\langle -, - \rangle$ on V by $\langle v, w \rangle = \varphi(v) \cdot \varphi(w)$, where the dot product on the right hand side happens in F^n.

Example 4.1.6. Set $V = C([a, b], \mathbb{C})$ and define $\langle f, g \rangle = \int_a^b f(t)\overline{g(t)}\, dt$. Then calculus says this makes V an inner product space. In $C([-\pi, \pi], \mathbb{C})$ with $f = 1 + 2x$ and $g = \cos(x)$, one can check that $\|f\| = \sqrt{2\pi + \frac{8\pi^3}{3}}$, $\|g\| = \sqrt{\pi}$, and that $\langle f, g \rangle = 0$.

Example 4.1.7. Let $V = M_n(\mathbb{C})$ and define $\langle A, B \rangle = \mathrm{tr}(B^*A)$, where $(A^*)_{ij} = \overline{A_{ji}}$ is the **conjugate transpose** of A. Then by linearity of tr and definition of B^*, one sees that this defines an inner product. For any $A \in M_n(\mathbb{C})$, we see that the ij-th entry of A^*A is simply the i-th row of A^* dotted with j-column of A. In particular, if v_i is the i-th column of A, then $(A^*A)_{ii} = \|v_i\|^2$, so that $\|A\| = \sqrt{\|v_1\|^2 + \ldots + \|v_n\|^2}$ where $v_1, \ldots, v_n$ are the columns of A. In $M_2(\mathbb{C})$, set $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} i & 0 \\ 1+i & -i \end{pmatrix}$. Then we see $\|A\| = \sqrt{30}$, $\|B\| = \sqrt{2}$, and $\langle A, B \rangle = 3$.

Example 4.1.8. Consider the sequence space $\mathbb{C}^\infty$. Let $\ell^2 = \{(a_n) \in \mathbb{C}^\infty : \sum_{n=1}^\infty |a_n|^2 < \infty\}$. For sequences $(a_n), (b_n)$, note that for any n, we have $2|a_n||b_n| \leq |a_n|^2 + |b_n|^2$. Combined with the triangle inequality, we then have $|a_n + b_n|^2 \leq (|a_n| + |b_n|)^2 \leq 2|a_n|^2 + 2|b_n|^2$, which then immediately tells us that ℓ^2 is a $\mathbb{C}$-vector space. Define $\langle (a_n), (b_n) \rangle = \sum_{n=1}^\infty a_n\overline{b_n}$. If this expression is finite, then it's clear that this makes ℓ^2 a complex inner product space, since it's just the usual dot product extended to vectors with infinitely many coordinates. To see this is the case, for any N, we have $\sum_{n=1}^N |a_n||\overline{b_n}| \leq \sqrt{\sum_{n=1}^N |a_n|^2 \sum_{n=1}^N |b_n|^2}$ by applying Cauchy-Schwarz to the vector space $\mathbb{R}^N$ with vectors $(|a_1|, \ldots, |a_N|)$ and $(|b_1|, \ldots, |b_N|)$ (here we use that for complex numbers, $|\bar{z}| = |z|$). Since $(a_n), (b_n) \in \ell^2$, taking $N \to \infty$ says the right hand side of the above inequality converges to something finite, so that $\lim_{N\to\infty} \sum_{n=1}^N |a_n||\overline{b_n}| = \sum_{n=1}^\infty |a_n||\overline{b_n}| < \infty$. This says $\sum_{n=1}^\infty a_n\overline{b_n}$ converges absolutely, so that $\sum_{n=1}^\infty a_n\overline{b_n}$ converges. We then have proved that this is indeed an inner product. This space is important in functional analysis.

4.2 Orthogonality

In $\mathbb{R}^n$, one of the most important properties of the dot product was that it was able to measure the angle between two vectors: this was detected by the quantity $\frac{x \cdot y}{\|x\|\|y\|}$. For a general inner product space V, it doesn't make sense to define general angles, since the expression $\frac{\langle x, y \rangle}{\|x\|\|y\|}$ may be a complex number. However, we may still make sense of orthogonality.

Definition 4.2.1. Vectors $x, y \in V$ are called **orthogonal** if $\langle x, y\rangle = 0$. A subset S of V is called orthogonal if any two vectors in S are orthogonal, and S is called **orthonormal** if S is orthogonal and $\|x\| = 1$ for all $x \in S$.

Since inner products have a notion of orthogonality, the Pythagorean theorem is still true:

Theorem 4.2.2 (Pythagorean theorem)**.** *Let $x, y \in V$ be orthogonal. Then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.*

Proof. $\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \langle x, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2$ because $\langle x, y\rangle = \langle y, x\rangle = 0$ by assumption. □

One of the reasons that we impose the extra structure of an inner product is that they become very nice to work with: orthogonality makes linear independence easy to check, as well as finding the coordinates of a vector with respect to some basis.

Proposition 37. *Let $\{v_1, \dots, v_k\}$ be an orthogonal subset of non-zero vectors. Then $\{v_1, \dots, v_k\}$ is linearly independent.*

Proof. If $c_1v_1 + \ldots + c_kv_k = 0$ is a linear dependence relation among the v_i's , then $\langle c_1v_1 + \ldots + c_kv_k, v_i\rangle = \langle 0, v_i\rangle = 0$. On the other hand, $\langle c_1v_1 + \ldots + c_kv_k, v_i\rangle = c_1\langle v_1, v_i\rangle + \ldots + c_k\langle v_k, v_i\rangle = c_i\|v_i\|^2$ by orthogonality, so $c_i = 0$. This says that $\{v_1, \dots, v_k\}$ is linearly independent. □

Proposition 38. *Let $S = \{v_1, \dots, v_k\}$ be an orthogonal subset of non-zero vectors. Then if $x = c_1v_1 + \ldots + c_kv_k$, $c_i = \frac{\langle x, v_i\rangle}{\|v_i\|^2}$.*

Proof. Taking an inner product with v_i says $\langle x, v_i\rangle = c_i\langle v_i, v_i\rangle = c_i\|v_i\|^2$ by orthogonality. □

In particular, the above says that if we have a basis β for V consisting of orthogonal vectors, then finding the coordinates $[x]_\beta$ is reduced to an inner product computation. If V is finite dimensional, is it always possible to find an orthonormal basis? The answer is yes, and follows from a more general result.

Theorem 4.2.3 (Gram-Schmidt process)**.** *Let $S = \{w_1, \dots, w_m\}$ be a linearly independent subset of V. Define $S' = \{v_1, \dots, v_m\}$ where $v_1 = w_1$ and $v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} v_j$ for $2 \leq j \leq m$. Then S' is an orthogonal subset of non-zero vectors and Span(S) = Span(S').*

Proof. The proof is by induction on m. If $m = 1$, then the result is immediate. Now suppose for every linearly independent set of size $m-1$ the theorem is true. Define $S_k = \{w_1, \dots, w_k\}$ and $S'_k = \{v_1, \dots, v_k\}$, so that in particular the set $S'_{m-1} = \{v_1, \dots, v_{m-1}\}$ is orthogonal. We will check that the theorem is true for $S' = S'_{m-1} \cup \{v_m\}$ where v_m is defined as in the statement of the theorem. If $v_m = 0$, this says $w_m \in \text{Span}(S'_{m-1})$, but $\text{Span}(S'_{m-1}) = \text{Span}(S_{m-1})$ by induction hypothesis, which contradicts that S is linearly independent. Therefore $v_m \neq 0$. We then see $\langle v_m, v_i\rangle = \langle w_m, v_i\rangle - \langle w_m, v_i\rangle = 0$ since by assumption S'_{m-1} is orthogonal, so that S' is therefore orthogonal. As $w_i \in \text{Span}(S'_{m-1})$ for all $1 \leq i \leq m-1$ by assumption, combined with the definition of v_m we get $w_i \in \text{Span}(S')$ for all i so that $\text{Span}(S) = \text{Span}(S')$ as desired. □

As an immediate corollary to the Gram-Schmidt process, we get the following:

Corollary 4.2.4. *If V is a finite dimensional inner product space, then V has an orthonormal basis.*

Proof. Apply the Gram-Schmidt process to a basis of V to get a basis of orthogonal vectors. Then normalize. $\square$

Example 4.2.5. Set $V = \mathbb{R}^3$ and $\beta = \{(1,1,1),(0,1,1),(0,0,1)\} = \{w_1, w_2, w_3\}$, which is a basis of $\mathbb{R}^3$. To construct an orthogonal basis, set $v_1 = (1,1,1)$. Then $v_2 = w_2 - \frac{\langle w_2, v_1\rangle}{\|v_1\|^2} v_1 = w_2 - \frac{2}{3}v_1 = (-2/3, 1/3, 1/3)$, and $v_3 = w_3 - \frac{\langle w_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2\rangle}{\|v_2\|^2} v_2 = w_3 - \frac{1}{3}v_1 - \frac{1}{2}v_2 = (0, -1/2, 1/2)$. This produces an orthogonal basis, so normalizing each vector with give an orthonormal basis. We see $\|v_1\| = \sqrt{3}$, $\|v_2\| = \sqrt{\frac{2}{3}}$, and $\|v_3\| = \frac{1}{\sqrt{2}}$. Then $\{(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}), (-\sqrt{\frac{2}{3}}, \sqrt{\frac{1}{6}}, \sqrt{\frac{1}{6}}), (0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})\}$ is an orthonormal basis of $\mathbb{R}^3$.

Example 4.2.6. Set $V = P_2(\mathbb{R})$,which may be viewed as a subspace of $C([-1,1])$ with the inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dt$. Let $\beta = \{1, x, x^2\} = \{w_1, w_2, w_3\}$ be the standard basis of V. To produce an orthonormal basis, we use Gram-Schmidt. The vectors 1 and x are already orthogonal, so we do not need to compute v_1 and v_2. Then $v_3 = w_3 - \frac{\langle w_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2\rangle}{\|v_2\|^2} v_2 = x^2 - \frac{1}{3}$, so $\{1, x, x^2 - \frac{1}{3}\}$ is an orthogonal basis. We compute $\|1\| = \sqrt{2}$, $\|x\| = \sqrt{\frac{2}{3}}$, and $\|x^2 - \frac{1}{3}\| = \sqrt{\frac{8}{45}}$. This produces an orthonormal basis $\{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1)\}$. These are the first three *Legendre polynomials*, which have applications in physics. Repeating this process with the basis $\beta = \{1, x, \ldots, x^n\}$ of $P_n(\mathbb{R})$ allows one to compute the n-th Legendre polynomial.

Definition 4.2.7. Let $S \subset V$ be a subset. The **orthogonal complement** of S, denoted $S^\perp$ is defined by $S^\perp = \{v \in V : \langle v, x\rangle = 0 \text{ for all } x \in S\}$.

It's an easy verification that $S^\perp$ is always a subspace of V. When S itself is a subspace, we have the following decomposition:

Theorem 4.2.8. *Let $W \subset V$ be a finite dimensional subspace. Then $V = W \oplus W^\perp$.*

Proof. Let $\{w_1, \ldots, w_k\}$ be an orthonormal basis of W. We will try to find a vector $w \in W$ such that $x = w + (x - w)$ with $x - w \in W^\perp$. Write $w = c_1w_1 + \ldots + c_kw_k$. If $x - w \in W^\perp$, then necessarily, $0 = \langle x - w, w_i\rangle = \langle x - c_1w_1 - \ldots - c_kw_k, w_i\rangle = \langle x, w_i\rangle - c_i\|w_i\|^2$. Since $\|w_i\| = 1$, this says $c_i = \langle x, w_i\rangle$, so this choice of coefficients gives us the vector w that works. This says $V = W + W^\perp$. If $w \in W \cap W^\perp$, then $\langle w, w\rangle = 0$ so that $w = 0$ says the sum is direct. $\square$

An immediate consequence is the following dimension formula:

Corollary 4.2.9. *If V is finite dimensional,* $\dim(V) = \dim(U) + \dim(U^\perp)$.

Using this, we easily get the following:

Proposition 39. *Let V be finite dimensional. Then $(W^\perp)^\perp = W$.*

Proof. Set $n = \dim(V)$. Since $W \subset (W^\perp)^\perp$, we get $\dim(W) \leq \dim(W^\perp)^\perp$. This then says $n = \dim(W) + \dim(W^\perp) \leq \dim((W^\perp)^\perp) + \dim(W^\perp) = n$ so that $\dim(W) = \dim((W^\perp)^\perp)$ gives $W = (W^\perp)^\perp$. □

Example 4.2.10. Let $V = \mathbb{R}^3$ and $W = \text{Span}\{v_1\}$ where $v_1 = (1,1,1)$. Then $W^\perp = \{(x,y,z) : (x,y,z) \cdot (1,1,1) = 0\}$, i.e. $W^\perp$ is simply the plane $x + y + z = 0$.

Example 4.2.11. Let $V = \mathbb{R}^4$ and $W = \text{Span}\{v_1, v_2\}$ where $v_1 = (1,2,3,-4)$ and $v_2 = (-5,4,3,2)\}$. If $x = (x,y,z,t)$ is in $U^\perp$, we see that $Ax = 0$, where $A = \begin{pmatrix} 1 & 2 & 3 & -4 \\ -5 & 4 & 3 & 2 \end{pmatrix}$. Using row reduction, one can easily compute $U^\perp = \ker(A) = \text{Span}\{(-3,-9,7,0),(-10,9,0,7)\}$.

Definition 4.2.12. Let W be a subspace with an orthonormal basis $\{w_1, \ldots, w_k\}$. Define the **orthogonal projection** onto W, P_W, by $P_W(x) = \langle x, w_1 \rangle w_1 + \ldots + \langle x, w_k \rangle w_k$.

Given $x \in V$, The orthogonal projection $P_W(x)$ has the property that it is the vector in W that is closest to x:

Theorem 4.2.13. *Let $x \in V$. Then $\|x - y\| \geq \|x - P_W(x)\|$ for all $y \in W$.*

Proof. Write $x = P_W(x) + z$ where $z \in W^\perp$. Then for any $y \in W$, we have $x - y = (P_W(x) - y) + z$. Then we see z is orthogonal to $P_W(x) - y$, so $\|x - y\|^2 = \|P_W(x) - y\|^2 + \|z\|^2 \geq \|P_W(x) - y\|^2$. □

Example 4.2.14. Let $V = \mathbb{R}^3$, and set $v = (1,2,3)$. What's the minimal distance from v to a point on the plane $W : x + 2y + z = 0$? A basis of W can be easily computed as $\{(-2,1,0),(-1,0,1)\} = \{w_1, w_2\}$. Running Gram-Schmidt gives an orthogonal basis of $\{v_1, v_2\} = \{(-2,1,0),(-5,-2,5)\}$. The minimal distance the the plane is given by the quantity $\|v - P_W(v)\|$. One can check that $P_W(v) = \frac{5}{3}v_2$, so $v - P_W(v) = (4/3, 8/3, 4/3)$ which has length $\frac{4\sqrt{6}}{3}$.

Example 4.2.15. Set $W = P_2(\mathbb{R})$ viewed as a subspace of $V = C([-1,1])$ with the inner product $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx$. With $f(x) = e^x$, which polynomial $p(x)$ of degree at most 2 minimizes the quantity $\int_{-1}^{1}(e^x - p(x))^2\,dt$, and what is this value? Equivalently, what is the minimizer of $\|e^x - p(x)\|$? We saw before than an orthonormal basis of $P_2(\mathbb{R})$ with respect to this inner product is given by the Legendre polynomials, with basis $\{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1)\}$, so the minimizer is just the orthogonal projection of e^x onto W. This is given by $p(x) = \langle e^x, \frac{1}{\sqrt{2}} \rangle \frac{1}{\sqrt{2}} + \langle e^x, \sqrt{\frac{3}{2}}x \rangle \sqrt{\frac{3}{2}}x + \langle e^x, \sqrt{\frac{5}{8}}(3x^2 - 1) \rangle \sqrt{\frac{5}{8}}(3x^2 - 1) = (\frac{15e}{4} - \frac{105}{4e})x^2 + \frac{3}{e}x + (\frac{33}{4e} - \frac{3e}{4})$. Numerically, the actual minimal value of the integral is $\approx .00144$.

4.3 The Adjoint of a Linear Operator

Definition 4.3.1. The **dual space** of V, denoted V^* is defined as $V^* = \text{Hom}_F(V, F)$. An element $\varphi \in V^*$ is called a **linear functional**.

If V is finite dimensional, then we have seen that $V \cong V^*$. However, this isomorphism is not "natural" in the sense that it requires picking a basis if V. However, when V is an inner product space, the isomorphism *is* natural:

Theorem 4.3.2 (Riesz Representation Theorem)**.** *Let V be a finite dimensional inner product space. Then the map $\Phi : V \to V^*$ given by $\Phi(v) = \varphi_v$ is an isomorphism, where $\varphi_v(x) = \langle x, v\rangle$.*

Proof. First, we show that Φ is linear. For $x, y \in V$, We have $\Phi(x + y) = \varphi_{x+y}$. For any $z \in V$, we have $\varphi_{x+y}(z) = \langle z, x+y\rangle = \langle z, x\rangle + \langle z, y\rangle = \varphi_x(z)+\varphi_y(z)$, so that $\varphi_{x+y} = \varphi_x+\varphi_y$. This then says that $\Phi(x + y) = \Phi(x) + \Phi(y)$. Similarly, we conclude that for any $c \in F$, $\Phi(cx) = c\Phi(x)$, so that Φ is linear. Now suppose that $\Phi(x) = 0$. This says that $\varphi_x(z) = 0$ for all $z \in V$, i.e. $\langle z, x\rangle = 0$ for all $z \in V$. Picking $z = x$, we get $\|x\|^2 = 0$, so that $x = 0$. This says that Φ is injective, and since $\dim V = \dim V^*$, we conclude that Φ is an isomorphism. □

The Riesz Representation Theorem says the structure of the dual space of a finite dimensional inner product space is very rigid: for any linear functional $\varphi \in V^*$, the surjectivity of the map Φ in the above proof says there is a vector $v \in V$ such that $\varphi = \langle -, v\rangle$. This is very important in functional analysis (where it holds in a more general setting), but for our purposes, we will only need it for the following:

Definition 4.3.3. Let V be a finite dimensional inner product space. The **adjoint** of a linear operator $T : V \to V$, denoted T^* is defined via the relation $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$.

Proposition 40. *The adjoint T^* of a linear operator T exists and is unique, and furthermore $T^* \in Hom_F(V, V)$.*

Proof. Define $\varphi_y(x) = \langle T(x), y\rangle$. Then $\varphi_y(x + z) = \langle T(x + z), y\rangle = \langle T(x) + T(z), y\rangle = \langle T(x), y\rangle + \langle T(z), y\rangle = \varphi_y(x) + \varphi_y(z)$. Similarly, $\varphi_y(cx) = c\varphi_y(x)$ for $x, z \in V$ and $c \in F$, so $\varphi_y(x)$ is a linear functional. By the Riesz Representation Theorem, $\varphi_y(x) = \langle x, y'\rangle$ for some $y' \in V$. Define a map $T^* : V \to V$ by $T^*(y) = y'$. By definition T^* satisfies the desired property. If there is another function $S : V \to V$ such that $\langle T(x), y\rangle = \langle x, S(y)\rangle$ for all x, y, this says $\langle x, T^*(y)\rangle = \langle x, S(y)\rangle$ for all x, y so that $T^* = S$. Finally, it remains to show linearity. We see $\langle x, T^*(y + z)\rangle = \langle T(x), y + z\rangle = \langle T(x), y\rangle + \langle T(x), z\rangle = \langle x, T^*(y)\rangle + \langle x, T^*(z)\rangle = \langle x, T^*(y) + T^*(z)\rangle$ for all $x, y, z \in V$. This says $T^*(y + z) = T^*(y) + T^*(z)$. Similarly one can check $T^*(cy) = cT^*(y)$, so that $T^* \in \mathrm{Hom}_F(V, V)$.

□

Although it may not be clear from the above defintion, the point of the adjoint is that it's a analogous operation on linear operators to taking a conjugate transpose. The following properties make this more clear:

Proposition 41. *Let $S, T \in Hom_F(V, V)$. The following hold:*

(a) $(S + T)^* = S^* + T^*$

(b) $(cT)^* = \overline{c}T^*$

(c) $(T^*)^* = T$

(d) $I^* = I$

(e) $(ST)^* = T^*S^*$

Proof. All the above properties can be proved using a similar approach to the one in the proposition above by pulling the adjoint through the inner product. We omit the proofs. □

Proposition 42. *Let V be a finite dimensional inner product space, and let β be an orthonormal basis of V. Then $[T^*]_\beta = [T]^*_\beta$.*

Proof. Let $\beta = \{v_1, \dots, v_n\}$ be an orthonormal basis for V. Set $[T]_\beta = [a_{ij}]$. Then $T(v_i) = a_{1i}v_1 + \ldots + a_{ni}v_n$, so $a_{ji} = \langle T(v_i), v_j\rangle$. This says $([T]^*_\beta)_{ij} = \overline{a_{ji}} = \overline{\langle T(v_i), v_j\rangle} = \langle v_j, T(v_i)\rangle = \langle T^*(v_j), v_i\rangle = ([T^*]_\beta)_{ij}$, so that $[T^*]_\beta = [T]^*_\beta$. □

Geometrically, the relationship between T^* and T is as follows:

Theorem 4.3.4. *Let V be a finite dimensional inner product space, and let $T : V \to V$ be a linear operator. Then* $\ker(T^*) = \operatorname{Im}(T)^\perp$ *and* $\operatorname{Im}(T^*) = \ker(T)^\perp$.

Proof. Let $x \in \ker(T^*)$, so that $T^*(x) = 0$. Then for any $y \in V$, $\langle y, T^*(x)\rangle = 0$. Pulling the adjoint through the inner product says $\langle T(y), x\rangle = 0$ for all y, so that $\ker(T^*) \subset \operatorname{Im}(T)^\perp$. Similarly, if $x \in \operatorname{Im}(T)^\perp$ this says $\langle x, T(y)\rangle = 0$ for all $y \in V$ so that $\langle T^*(x), y\rangle = 0$ for all $y \in V$. This says $T^*(x) = 0$, so that $\operatorname{Im}(T)^\perp \subset \ker(T^*)$ says $\ker(T^*) = \operatorname{Im}(T)^\perp$. Setting $T = T^*$ and taking orthogonal complements of both sides gives the second statement. □

Example 4.3.5. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be given by $T(z_1, z_2) = (z_1 - 2iz_2, 3z_1 + iz_2)$, where $\mathbb{C}^2$ is equipped with the usual dot product. Then the standard basis $\{e_1, e_2\}$ is orthonormal. We see $[T]_\beta = \begin{pmatrix} 1 & -2i \\ 3 & i \end{pmatrix}$, so that $[T^*]_\beta = [T]^*_\beta = \begin{pmatrix} 1 & 3 \\ 2i & -i \end{pmatrix}$.

Example 4.3.6. Let $T : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ be the transpose map, $T(A) = A^t$. Equip $M_2(\mathbb{R})$ with the inner product $\langle A, B\rangle = \operatorname{tr}(B^tA)$. With respect to this inner product, the standard basis $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ is orthonormal. Then $[T^*]_\beta = [T]^*_\beta = [T]^t_\beta$. We see that $[T]_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. This matrix is symmetric, so $T = T^*$.

Example 4.3.7. Let $V \subset C^\infty(\mathbb{R})$ be the vector space of infinitely differentiable functions that are 1-periodic, i.e. $f(x+1) = f(x)$ for all $x \in \mathbb{R}$. Give V an inner product structure by $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$. Let $D : V \to V$ be the derivative map. To compute the adjoint of D, we use the definition. For $f, g \in V$, $\langle D(f), g\rangle = \int_0^1 f'(t)g(t)\,dt$. Integrating by parts and using $f(1) = f(0)$, the latter integral equals $-\int_0^1 f(t)g'(t)\,dt = \langle f(t), -D(g)\rangle$. This says $D^* = -D$.

4.4 The Spectral Theorem

We will now return to diagonalizability. We previously saw what conditions are necessary for a linear operator on V to be diagonalizable, i.e. for V to have a basis of eigenvectors for T. If V is an inner product space, a natural question is when can we find an *orthonormal* basis of eigenvectors? The Spectral theorem gives a precise answer.

Definition 4.4.1. A linear operator $T : V \to V$ is called **normal** if $TT^* = T^*T$. T is called **self-adjoint** if $T = T^*$.

Proposition 43. *Suppose that $T : V \to V$ is normal. Then if v is an eigenvector of T with eigenvalue λ, then v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$.*

Proof. It's easy to check that since T is normal, then so is $T - cI_V$ for any $c \in F$. Since $T(v) = \lambda v$, this says $0 = \|(T-\lambda I_V)(v)\|^2 = \langle (T-\lambda I_V)(v), (T-\lambda I_V)(v)\rangle = \langle v, (T^*-\overline{\lambda} I_V)(T-\lambda I_V)(v)\rangle = \langle v, (T-\lambda I_V)(T^*-\overline{\lambda} I_V)(v)\rangle = \langle (T^*-\overline{\lambda} I_V)(v), (T^*-\overline{\lambda} I_V)(v) = \|(T^*-\overline{\lambda} I_V)(v)\|^2$. This says $T^*(v) = \overline{\lambda} v$ as desired. □

Proposition 44. *Suppose that $T : V \to V$ is normal. Then if λ_1, λ_2 are distinct eigenvalues of T with eigenvectors v_1 and v_2 respectively, then v_1 and v_2 are orthogonal.*

Proof. Suppose $T(v_1) = \lambda_1 v_1$ and $T(v_2) = \lambda_2 v_2$. Then $\langle T(v_1), v_2\rangle = \lambda_1 \langle v_1, v_2\rangle$. On the other hand, $\langle T(v_1), v_2\rangle = \langle v_1, T^*(v_2)\rangle = \langle v_1, \overline{\lambda_2} v_2\rangle = \lambda_2\langle v_1, v_2\rangle$ by the above proposition. Since $\lambda_1 \neq \lambda_2$, this forces $\langle v_1, v_2\rangle = 0$. □

Theorem 4.4.2 (Complex Spectral Theorem)**.** *Let V be a finite dimensional complex inner product space. Then a linear operator $T : V \to V$ is normal if and only if there is an orthonormal basis for V consisting of eigenvectors of T.*

Proof. First suppose that T is normal. We prove that T is orthogonally diagonalizable by induction on the dimension of V. If $\dim(V) = 1$ then this is obvious, because any non-zero vector is an eigenvector, so just normalize. Now suppose that any normal operator on an $n-1$ dimensional complex inner product space is orthogonally diagonalizable. If $\dim(V) = n$ and $T : V \to V$ is a normal operator, because $\mathbb{C}$ is algebraically closed T has an eigenvector, say v. Set $U = \operatorname{Span}(\{v\})$ and write $V = U \oplus U^\perp$. Note that because T is normal, both T, T^* are U-invariant. If $x \in U^\perp$, then for $y \in U$, we have $\langle y, T(x)\rangle = \langle T^*(y), x\rangle = 0$ because $T^*(y) \in U$. This says $T(x) \in U^\perp$ so that T is $U^\perp$-invariant. Similarly, T^* is $U^\perp$-invariant. Then we may write $T(x) = T|_U(u) + T|_{U^\perp}(u')$ for $x = u + u'$ with $u \in U$ and $u' \in U^\perp$. We now show that $T|_{U^\perp}$ is a normal operator on $U^\perp$.

By definition, for $x, y \in U^\perp$, $\langle T|_{U^\perp}(x), y\rangle = \langle x, (T|_{U^\perp})^*(y)\rangle$. However, by definition $T|_{U^\perp}$ and T agree on $U^\perp$, so $\langle T|_{U^\perp}(x), y\rangle = \langle T(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, (T^*)|_{U^\perp}(y)\rangle$. This says $(T|_{U^\perp})^* = (T^*)_{U^\perp}$. Then $(T|_{U^\perp})(T|_{U^\perp})^* = T|_{U^\perp}(T^*)|_{U^\perp} = TT^* = T^*T = (T^*)|_{U^\perp}T|_{U^\perp} = (T|_{U^\perp})^*(T|_{U^\perp})$, which proves $T|_{U^\perp}$ is normal. By induction, there is an orthonormal basis $\{v_2, \ldots, v_n\}$ of $U^\perp$ consisting of eigenvectors for $T|_{U^\perp}$. Then $\{v, v_2, \ldots, v_n\}$ is an orthogonal basis of V consisting of eigenvectors of T. Normalizing v makes this orthonormal, so we are done.

Conversely, suppose that T is orthogonally diagonalizable. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of eigenvectors with eigenvalues λ_i. Then $(T^*T)(v_i) = T^*(\lambda_i v_i) = \lambda_i T^*(v_i) = |\lambda_i|^2 v_i$. On the other hand, $(TT^*)(v_i) = T(\overline{\lambda_i} v_i) = |\lambda_i|^2 v_i$. Then T^*T and TT^* agree on a basis of V, so they are equal which shows T is normal as desired. □

We now move onto the Spectral Theorem for operators on real inner product spaces. In the complex case, we were able to make the argument work because the fundamental theorem of algebra says every linear operator over a complex vector space has an eigenvalue, which led to a decomposition $V = U \oplus U^\perp$. The key part of the proof is the normality of T said that it restricted to *normal* operators on U and $U^\perp$, allowing the induction to kick in. If V is a real inner product space, this no longer remains true, as we have seen that a rotation by some angle in $\mathbb{R}^2$ has no *real* eigenvalue. If we can find a class of normal operators that are guaranteed to have a real eigenvalue, then the same argument as above goes through. As it turns out, the key to this is self-adjointness:

Proposition 45. *Suppose that $T : V \to V$ is self-adjoint. Then if λ is an eigenvalue of T, then λ is real.*

Proof. Write $T(v) = \lambda v$. Then $\langle T(v), v\rangle = \lambda\langle v, v\rangle = \lambda\|v\|^2$. On the other hand, because T is self-adjoint we can write $\langle T(v), v\rangle = \langle v, T(v)\rangle = \overline{\lambda}\|v\|^2$. Since v is non-zero, this says $\lambda = \overline{\lambda}$ so that λ is real. □

Theorem 4.4.3. *(Real Spectral Theorem) Let V be a finite dimensional real inner product space. Then a linear operator $T : V \to V$ is self-adjoint if and only if there is an orthonormal basis for V consisting of eigenvectors of T.*

Proof. The characteristic polynomial p_T of T has a complex root by the fundamental theorem of algebra. Since T is self-adjoint, the above says this root is real, so that T has an eigenvector. Since a self-adjoint operator is normal, we can run the same argument in the complex case and the proof still goes through, so that T is orthogonally diagonalizable.

Conversely, suppose that T is orthogonally diagonalizable. The argument from before shows that T is normal. Let $\beta = \{v_1, \ldots, v_n\}$ be an eigenbasis for V with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $T(v_i) = \lambda_i v_i$, and $T^*(v_i) = \overline{\lambda_i} v_i$. However, λ_i are *real*, which says that $T = T^*$ so that T is self-adjoint. □

We then immediately get the corresponding statements for matrices:

Corollary 4.4.4. *Let $T : V \to V$ be a normal operator over a finite dimensional complex inner product space, or a self-adjoint operator on a real inner product space. Then there is an orthonormal basis γ of V such that $[T]_\gamma = PDP^*$ where P is orthogonal, i.e. $PP^* = P^*P = I$ and D is diagonal.*

Proof. Fix an orthonormal basis β of V. By the Spectral Theorem, there is a basis γ of V consisting of orthonormal eigenvectors of T. Then $S^{\beta}_{\beta'}$ is orthogonal, so the change of basis formula gives the result with $P = S^{\beta}_{\beta'}$ and D the diagonal matrix of eigenvalues of T. □

The proof of the Spectral Theorem tells us how to orthogonally diagonalize an operator when it is possible. If $V = U \oplus U^{\perp}$, running Gram-Schmidt on bases of U and $U^{\perp}$ give orthogonal bases of these spaces, and then the union is an orthogonal basis of V, so after normalizing, an orthonormal basis. Suppose T is normal with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. In the proof of the Spectral Theorem, we may instead run the argument with $U = E_{\lambda_1}$ (the invariance condition is still true). Then since $E_{\lambda_i} \perp E_{\lambda_1}$ for $i \neq 1$, this says $E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k} \subset U^{\perp}$ so that $E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k} = U^{\perp}$ for dimensional reasons. By inductively applying the above obeservation, this says running Gram-Schmidt on each eigenspace E_{λ_i} and taking the union of these orthogonal basis is then an orthogonal basis for V consisting of eigenvalues of T, and then normalizing gives an orthonormal basis.

Example 4.4.5. The operator $T : \mathbb{C}^2 \to \mathbb{C}^2$ given by $T(z_1, z_2) = (z_2, 0)$ is not normal, because it is not diagonalizable.

Example 4.4.6. The operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (-y, x)$ is normal. With respect the the standard basis, $[T]_\beta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, so $[T^*]_\beta = [T]^t_\beta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -[T]_\beta$. However, T is not self-adjoint, because $[T]_\beta$ is not a symmetric matrix. T has no real eigenvalues so it is not diagonalizable over $\mathbb{R}$, but over $\mathbb{C}$ has eigenvalues $i, -i$. To orthogonally diagonalize T over $\mathbb{C}$, a basis of eigenvectors is given by $\{(i, 1), (-i, 1)\}$, which we see is orthogonal. Normalizing says an orthonormal basis of eigenvectors is $\beta' = \{(\frac{i}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (\frac{-i}{\sqrt{2}}, \frac{1}{\sqrt{2}})\}$. Since β' is orthonormal, the change of basis matrix $S^{\beta}_{\beta'}$ satisfies the relation $(S^{\beta}_{\beta'})^{-1} = S^{\beta'}_{\beta} = (S^{\beta}_{\beta'})^*$ This gives the matrix factorization $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{i}{\sqrt{2}} & \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} \frac{-i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.

Example 4.4.7. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (-2z, -x + 2y - z, x + 3z)$, so that $[T]_\beta = \begin{pmatrix} 0 & 0 & -2 \\ -1 & 2 & -1 \\ 1 & 0 & 3 \end{pmatrix}$ with β the standard basis. We see that T is diagonalizable with eigenvalues $1, 2$ and basis of the eigenspaces E_1 and E_2 are given by $\{(2, 1, -1)\}$ and $\{(0, 1, 0), (-1, -1, 1)\}$ respectively. However, T is not self-adjoint because $[T]_\beta$ is not symmetric, so the Spectral Theorem says that T is not orthogonally diagonalizable. What goes wrong? An orthogonal basis of E_2 is given by $\{(0, 1, 0), (1, 0, -1)\}$. However, $(2, 1, -1) \cdot (0, 1, 0) = 1 \neq 0$. Since any eigenvector $v \in E_2$ is of the form $(c_2, c_1, -c_2)$ for $c_1, c_2 \in \mathbb{R}$, we see that $(2, 1, -1) \cdot (c_2, c_1, -c_2) = 2c_1 + 2c_2$ is 0 only when $c_2 = -c_1$, i.e. the eigenvector is of the form $(-c_1, c_1, c_1)$. Therefore it's impossible to find two eigenvectors orthogonal to $(2, 1, -1)$, so that T cannot be orthogonally diagonalizable. Explicitly, with $U = E_1$, we see that $[T^*]_\beta = [T]^t_\beta$. T^* is not U-invariant, because $T^*(2, 1, -1) = (-2, 2, -8) \notin U$, so that T^* is not U-invariant and the argument cannot continue. Since all the eigenvalues of T are real, we see that even viewed as an operator on $\mathbb{C}^3$, the only eigenvector in E_2 that is orthogonal to $(2, 1, -1)$ is in the $\mathbb{C}$-span of $(-1, 1, 1)$, so again it is not possible to find two eigenvectors orthogonal to $(2, 1, -1)$. This then says that T is not normal when viewed as an operator on $\mathbb{C}^3$, and therefore not as an operator on $\mathbb{R}^3$ because the matrix of T^* is the same in either case.