LINEAR ALGEBRA

MICHAEL PENKAVA

1. Linear Maps

Definition 1.1. If $V$ and $W$ are vector spaces over the same field $K$, then a map $\lambda : V \to W$ is called a linear map if it satisfies the two conditions below:

1. $\lambda(v_1 + v_2) = \lambda(v_1) + \lambda(v_2)$. (Additivity)
2. $\lambda(rv) = r\lambda(v)$. (Homogeneity)

A linear map is also called a linear transformation. A linear map $T : V \to V$, in other words, one from a vector space to itself, is often called a linear operator, although some authors use this terminology for an arbitrary linear map.

Linear maps are the main object of study in linear algebra. Many examples of linear maps come from calculus, so it is not surprising that a linear algebra course usually expects a calculus prerequisite, even though the concept of linear algebra does not require any calculus knowledge.

Example 1.2. Let $V = C[a,b]$ be the vector space of real valued continuous functions on the interval $[a,b]$. Then the map $I : V \to \mathbb{R}$, given by

$$I(f) = \int_{a}^{b} f(x) \, dx$$

is a linear map. The additivity of the integral is the rule that the integral of a sum of functions is the sum of the integrals. The homogeneity condition is the constant coefficient rule for integrals.

The linear property of integrals leads to many similar linear maps.

Example 1.3. Let $C^n(X)$ be the space of functions on an open set $X$ in $\mathbb{R}$ which satisfy the property that they are $n$-times differentiable and their $n$-th derivative is continuous, that is, is an element of the space $C^n(X)$ of continuous functions on $X$. Then the map $D : C^n(X) \to C^{n-1}(X)$ given by $D(f)(x) = f'(x)$ is a linear map for $n \geq 1$, by the sum and constant multiple rules of the derivative.

Again, like for integration, there are many variants of this type of linear map. For example, if $C^\infty(X)$ is the space of functions on $X$ which have derivatives of all orders, then $D : C^\infty(X) \to C^\infty(X)$ is a linear map.

Example 1.4. A basic example of a linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$. Define the map $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x_1, x_2, x_3) = (3x_1 - x_3, 5x_1 + 2x_2 - 4x_3)$. To show that $T$ is
Define a map

We have

Let

Suppose that

shows immediately that

Example 1.8.

M

vectors

is a linear transformation is immediate. Moreover, this example

Example 1.5. Define a map

T

by a property relating matrix and scalar multiplication. Thus

It is easy to check that matrix multiplication always gives a linear transformation, so the fact that

The identity map is often denoted as

vector space

is easy to check that 0 is a linear map. Also, there is a special linear map between a

Given two \( \mathbb{K} \)-vector spaces \( V \) and \( W \) there is always a linear transformation between them, called the \textit{trivial map} \( 0 : V \to W \), which is defined by \( 0(v) = 0 \). It is easy to check that \( 0 \) is a linear map. Also, there is a special linear map between a vector space \( V \) called the \textit{identity map} \( I : V \to V \), given by \( I(v) = v \) for all \( v \in V \).

The identity map is often denoted as \( I_V \) or \( 1_V \), to emphasize the vector space \( V \).

To verify that a map \( T : V \to W \) is not a linear transformation, one can find vectors \( v, v' \in V \) such that \( T(v + v') \neq T(v) + T(v) \) or find a scalar \( c \) and a vector \( v \in V \) such that \( T(cv) \neq cT(v) \). A very simple method of showing that a map is not linear is given by applying the contrapositive to the following simple proposition.

Proposition 1.7. Suppose that \( T : V \to W \) is linear. Then \( T(0) = 0 \).

Example 1.8. Let us show that the map \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) given by \( T(x, y) = (x + 1, 2x + y) \) is not linear. Computing, we obtain \( T(0) = T(0,0) = (1,0) \neq 0 \). This shows immediately that \( T \) is not linear.
Theorem 1.9. Let \( T : V \to W \) be a linear map of two \( \mathbb{K} \)-vector spaces. Then
\[
T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n)
\]
for any \( n \in \mathbb{N} \), \( a_1, \ldots, a_n \in \mathbb{K} \) and \( v_1, \ldots, v_n \in V \).

Exercise 1.10. Use induction on \( n \in \mathbb{N} \) to prove the theorem above.

The theorem above has a very important application to constructing linear transformations from bases of vector spaces. We state it as a theorem.

Theorem 1.11. Let \( V = \langle v_1, \ldots \rangle \) be a basis of a \( \mathbb{K} \)-vector space \( V \), and suppose \( w_1, \ldots \) are elements of a \( \mathbb{K} \)-vector space \( W \). Then there is a unique linear transformation \( T : V \to W \) such that \( T(v_i) = w_i \) for \( i = 1, \ldots \). Moreover, if \( v = a_1v_1 + \cdots + a_nv_n \), then
\[
T(v) = a_1w_1 + \cdots + a_nw_n.
\]

Proof. Define \( T \) as follows. Let \( v \in V \). Then \( v \) has a unique expression in the form \( v = a_1v_1 + \cdots + a_nv_n \) for some coefficients \( a_i \). Let \( T(v) = a_1w_1 + \cdots + a_nw_n \). Let us show that \( T \) is linear. Suppose that \( v' = a_1'v_1 + \cdots + a'_nv_n \). Then
\[
T(v + v') = T((a_1 + a'_1)v_1 + \cdots + (a_n + a'_n)v_n)
= (a_1 + a'_1)w_1 + \cdots + (a_n + a'_n)w_n
= a_1w_1 + \cdots + a_nw_n + a'_1w_1 + \cdots + a'_nw_n
= T(v) + T(v').
\]
\[
T(cv) = T(c_1v_1 + \cdots + c_nv_n) = c_1a_1w_1 + \cdots + c_nw_n
= c(a_1w_1 + \cdots + a_nw_n)
= cT(v).
\]

This shows that there is at least one linear transformation \( T : V \to W \) satisfying \( T(v_i) = w_i \). On the other hand, if \( T \) is any such linear transformation, then by the theorem above, it must satisfy \( T(v) = a_1T(v_1) + \cdots + a_nT(v_n) = a_1w_1 + \cdots + a_nw_n \), so the formula for \( T \) above defines \( T \) uniquely.

Note that in the theorem above, we really don’t need the vector spaces \( V \) or \( W \) to be finite dimensional. The only thing we need is a basis of \( V \). In an advanced linear algebra class you will learn that every vector space has a basis, using Zorn’s lemma, which is an axiom of set theory.

Example 1.12. Let \( \langle 1, x, \ldots \rangle \) be the standard basis of \( \mathbb{K}[x] \). Define \( T(x^n) = \int_0^\infty x^n \exp(-x) \, dx \). Then \( T \) extends uniquely to a linear map \( T : \mathbb{K}[x] \to \mathbb{R} \). In fact, we can determine \( T \) explicitly in several manners. First, we have \( T(p(x)) = \int_0^\infty p(x) \, dx \) which is an expression for \( T \) that does not depend on the basis. Secondly, we actually can compute the integral \( \int x^n \exp(-x) \, dx = n! \) using calculus, so that we can calculate
\[
T(a_nx^n + \cdots + a_0) = \sum_{k=0}^n a_k k!.
\]

The space \( \mathbb{K}[x] \) is infinite dimensional, but we still could use the construction in the theorem above to define a linear transformation using a basis of \( \mathbb{K}[x] \).
Let $T : V \to W$ be a linear transformation. Then the kernel of $T$, denoted as $\ker(T)$, is the set

$$\ker(T) = \{ v \in V | T(v) = 0 \}.$$ 

The nullity of $T$, denoted $\text{nullity}(T)$, is the dimension of the kernel of $T$ and the rank of $T$, denoted $\text{rank}(T)$, is the dimension of the image of $T$.

The image of $T$, denoted $\text{Im}(T)$, is the subset of $W$ given by

$$\text{Im}(T) = \{ T(v) | v \in V \}.$$ 

The image of a linear transformation is often called the range of the linear transformation. However, some authors mean the target space $W$ as the range of $T$, instead of the image. Thus it is perhaps better to avoid the use of the term range, as it can have multiple meanings. However, we shall use the terms image and range interchangeably in this document, and refer to the space $W$ as the target space.

Proposition 1.14. Let $T : V \to W$ be a linear transformation. Then the kernel of $T$ is a subspace of $V$ and the image of $T$ is a subspace of $W$.

Exercise 1.15. Prove the theorem above.

Lemma 1.16. Let $T : V \to W$ be a linear transformation and $\langle w_1, \ldots, w_k \rangle$ be linearly independent elements in $\text{Im}(T)$. Suppose that $v_1, \ldots, v_k$ are any vectors in $T$ such that $T(v_i) = w_i$ for $i = 1, \ldots, k$. Then $S = \{ v_1, \ldots, v_k \}$ is a linearly independent set.

Proof. Suppose that $a_1 v_1 + \cdots + a_k v_k = 0$. Then

$$0 = T(0) = T(a_1 v_1 + \cdots + a_k v_k) = a_1 T(v_1) + \cdots + a_k T(v_k) = a_1 w_1 + \cdots + a_k w_k.$$ 

Since $w_1, \ldots, w_k$ is linearly independent, it follows that $a_1 = \cdots = a_k = 0$. But this shows that $v_1, \ldots, v_k$ is linearly independent. \qed

Theorem 1.17. Let $T : V \to W$ be a linear map. Then

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

Proof. If $\text{nullity}(T) = \infty$, then $\dim(V) = \infty$, since the nullity of $T$ is the dimension of the nullspace of $T$, which is a subspace of $V$. Thus the statement of the theorem holds, since both sides are equal to infinity. If $\text{rank}(T) = \infty$, then the dimension of the image of $T$ is infinity, so we can find an infinite sequence of elements $w_1, \ldots, w_\infty$ in $\text{Im}(T)$ which are linearly independent. Since each of these elements is in the image of $T$, we can find an infinite sequence of elements $v_1, \ldots$ such that $T(v_i) = w_i$. By the lemma, this sequence is linearly independent, so is a basis for an infinite dimensional subspace of $V$. In this case, we also see that $\dim(V) = \infty$, so the equation holds.

Now we can assume that both nullity($T$) and rank($T$) are finite. Let $\langle w_1, \ldots, w_m \rangle$ be a basis of $\text{Im}(T)$ so that $\text{rank}(T) = m$. Let $v_1, \ldots, v_m$ be chosen so that $T(v_i) = w_i$ for $i = 1, \ldots, m$. Since $v_1, \ldots, v_m$ are a linearly independent set, they are the basis of a subspace $U$ of $V$ of dimension $m$. We claim that

$$V = \ker(T) \oplus U.$$ 

To see this, suppose that $v \in V$. Since $w_1, \ldots, w_m$ is a basis of the image of $T$, $T(v) = a_1 w_1 + \cdots + a_m w_m$ for some $a_1, \ldots, a_m \in \mathbb{K}$. Let $u = a_1 v_1 + \cdots + a_m v_m$ so that $T(u) = T(v)$. But then $v = v - u + u$ and $T(v - u) = T(v) - T(u) = 0$, so
Find a basis of the kernel and image in example (1.4). Suppose $A$ is a map. Then the following are equivalent:

- $T$ is injective
- $T$ is surjective
- $T$ is an isomorphism

Proof. Suppose $\dim(V) = \dim(W) < \infty$ and $T : V \to W$ is a linear map. If $T$ is injective, then $\ker(T) = \{0\}$. Thus $\dim(\ker(T)) = 0$ so $\dim(V) = \dim(\ker(T))$. Thus $\dim(V) = \dim(W)$, so $T$ is surjective. Thus (1) implies (2).

Similarly, if $T$ is surjective, then $\dim(V) = \dim(W) = \dim(V)$, which forces $\ker(T) = \{0\}$. Thus $T$ is injective. Thus (2) implies (3).

Finally, if $T$ is an isomorphism, then it is injective, so (3) implies (1).
Remark 1.23. When $V$ and $W$ are not finite dimensional, the above theorem is no longer true. For example, the map $T : \mathbb{K}[x] \to \mathbb{K}[x]$ given by $T(x^n) = x^{2n}$ is injective, but not surjective, while the map $S : \mathbb{K}[x] \to \mathbb{K}[x]$ given by $S(x^{2n}) = x^n$ and $S(x^{2n+1}) = 0$ is surjective, but injective.

Proposition 1.24. Let $T : V \to W$ be a linear map. If $W$ is finite dimensional, then $T$ is surjective if and only if $\text{rank}(T) = \text{dim}(W)$. 

Proof. Exercise. 

Definition 1.25. Let $V$ and $W$ be $\mathbb{K}$-vector spaces. Then we say that $V$ and $W$ are isomorphic, and write $V \cong W$, if there is an isomorphism $T : V \to W$.

Theorem 1.26. Let $V$ and $W$ be finite dimensional $\mathbb{K}$-vector spaces. Then $V$ and $W$ are isomorphic if and only if $\text{dim}(V) = \text{dim}(W)$.

Proof. Suppose that $V$ and $W$ are isomorphic. Let $T : V \to W$ be an isomorphism, and suppose $V = \langle v_1, \ldots, v_n \rangle$. Then $\text{dim}(V) = n$. Let $w_i = T(v_i)$ for $i = 1, \ldots, n$. Then we claim that $W = \langle w_1, \ldots, w_n \rangle$. Let $w \in W$ and suppose that $v \in V$ is such that $T(v) = w$. Now $v = a_1v_1 + \cdots + a_nv_n$ for some $a_1, \ldots, a_n \in \mathbb{K}$, because $v_1, \ldots, v_n$ is a basis of $V$. Thus $w = T(v) = T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n) = a_1w_1 + \cdots + a_nw_n$, so lies in the span of $w_1, \ldots, w_n$. On the other hand, suppose that $a_1w_1 + \cdots + a_nw_n = 0$. But then $0 = a_1w_1 + \cdots + a_nw_n = a_1T(v_1) + \cdots + a_nT(v_n) = T(a_1v_1 + \cdots + a_nv_n)$. 

Since $T$ is an isomorphism, it follows that $a_1v_1 + \cdots + a_nv_n = 0$. But $v_1, \ldots, v_n$ is a basis of $V$, so it follows that $a_1 = \cdots a_n = 0$. This shows that $w_1, \ldots, w_n$ is a linearly independent set of elements in $W$ which span $W$. It follows that $\text{dim}(W) = n = \text{dim}(V)$.

On the other hand, suppose that $\text{dim}(V) = \text{dim}(W) = n$. Then $V = \langle v_1, \ldots, v_n \rangle$ for some set $v_1, \ldots, v_n$ and $W = \langle w_1, \ldots, w_n \rangle$ for some set $w_1, \ldots, w_n$. Let $T$ be the unique linear transformation such that $T(v_i) = w_i$ for $i = 1, \ldots, n$. Then we claim that $T$ is an isomorphism between $V$ and $W$, so that $V$ and $W$ are isomorphic. First, note that if $T(v) = 0$, then, since $v = a_1v_1 + \cdots + a_nv_n$, we have $0 = T(v) = T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n) = a_1w_1 + \cdots + a_nw_n$, which forces $a_1 = \cdots a_n = 0$, because $w_1, \ldots, w_n$ is a basis of $W$. But this means that $v = 0$, so we see that $\ker(T) = 0$ and $T$ is injective. On the other hand, if $w \in W$, then $w = a_1w_1 + \cdots + a_nw_n$ for some $a_1, \ldots, a_n \in \mathbb{K}$. It follows that $w = T(v)$ where $v = a_1v_1 + \cdots + a_nv_n$. Thus $T$ is surjective. Since we have show that $T$ is injective and surjective, we see that it is an isomorphism. Thus $V$ and $W$ are isomorphic.

Example 1.27. The spaces $\mathbb{K}^n$ and $\mathbb{M}_{n,1}(\mathbb{K})$ are isomorphic because they both have dimension $n$. Both $\mathbb{R}^n$ and $\mathbb{M}_{n,1}$ have standard bases. The standard basis of $\mathbb{R}^n$ $S = e_1, \ldots, e_n$, where $e_i$ is the element of $\mathbb{K}^n$ which has a 1 in the $i$-th spot and zeros elsewhere. In other words,

$$e_1 = (1,0,\ldots,0), \quad e_2 = (0,1,0,\ldots,0), \ldots, e_n = (0,\ldots,0,1).$$

The standard basis $S'$ of $\mathbb{M}_{n,1}$ is given by $e'_{ij} = e_{ij}$ where $e_i$ is the standard basis of $\mathbb{M}_{m,n}$. The canonical isomorphism $T : \mathbb{R}^n \to \mathbb{M}_{n,1}$ is the isomorphism given by
Whenever two vector spaces of the same dimension have standard bases, the isomorphism defined in this way is called the canonical or natural isomorphism. Of course, there are many isomorphisms between vector spaces of the same dimension, but they are not considered as canonical unless the spaces are equipped with standard bases.

**Theorem 1.28.** Let \( B \) be an ordered basis of \( V \), \( C \) be an ordered basis of \( W \) and \( T : V \to W \) be a linear map. Suppose that both \( V \) and \( W \) are finite dimensional. If \( B = \{e_1, \ldots, e_n\} \), Define the matrix \( T_{C,B} \) by

\[
T_{C,B} = [T(e_1)]_C | \cdots | [T(e_n)]_C.
\]

Then for any \( v \in V \), we have

\[
T_{C,B}[v]_B = [T(v)]_C.
\]

**Proof.** Recall that \( v|_B \) is the coordinate vector of \( v \) with respect to the basis \( B \), which means that \( [v]_B = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \), where \( v = x_1 e_1 + \cdots + x_n e_n \). We compute

\[
T_{C,B}[v]_B = [T(e_1)]_C | \cdots | [T(e_n)]_C \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 [T(e_1)]_C + \cdots + x_n [T(e_n)]_C = [x_1 T(e_1)]_C + \cdots + [x_n T(e_n)]_C = [T(x_1 e_1 + \cdots + x_n e_n)]_C = [T(v)]_C.
\]

The matrix \( T_{C,B} \) is called the matrix of \( T \) with respect to the input basis \( B \) and output basis \( C \). One of the advantages of studying the matrix of a linear transformation (with respect to some input and output basis) is that it reduces the problem of computing the kernel of a linear transformation to the computation of the nullspace of the matrix, and the problem of finding a basis of the image to finding a basis of the column space of the matrix. These are problems which you already know how to solve by using Gaussian or Gauss-Jordan elimination of the matrix.

**Example 1.29.** Consider the linear transformation in example (1.4). In example (1.5) we computed the matrix of this linear transformation with respect to the standard bases \( S' = \{e_1, e_2, e_3\} \) of \( \mathbb{R}^3 \) and \( S = \{e_1, e_2\} \) of \( \mathbb{R}^2 \). Let us verify this by a direct calculation. \( T(e_1) = T(1,0,0) = (3,5) \), so \( [T(e_1)]_S = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \). \( T(e_1) = T(0,1,0) = (0,2) \), so \( [T(e_2)]_S = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \). Finally, \( T(e_3) = T(0,0,1) = (-1,-4), \) so \( [T(e_3)]_S = [-1,-4] \). Thus, we obtain \( T_{S,S} = \begin{bmatrix} 3 & 0 & -1 \\ 5 & 2 & -4 \end{bmatrix} \), which is the matrix we obtained in example (1.5).

**Example 1.30.** In practise, it is easy to write down the matrix of a linear transformation with respect to standard bases of the input and output spaces, as long as
the map is defined in terms of these bases. For example, let $S$ be the standard basis of $\mathbb{R}^3$ and $S' = \{1, x, x^2, x^3\}$ be the standard basis of $P_3(\mathbb{R})$. Then if

$$T(a_3x^3 + a_2x^2 + a_1x + a_0) = (a_3 - 2a_2 + 5a_1 + 3a_0, -2a_3 + a_1 - 2a_0, -4a_3 + a_2 - a_1 + a_0),$$

Then

$$T_{S,S'} = \begin{bmatrix} 3 & 5 & -2 & 1 \\ -2 & 1 & 0 & -2 \\ 1 & -1 & 1 & -4 \end{bmatrix}. $$

**Theorem 1.31.** Let $T : U \to V$ and $S : V \to W$ be linear transformations of $\mathbb{K}$-vector spaces. Then $S \circ T : U \to W$ is also a linear transformation.

**Proof.** Let $u, u' \in U$ and $c \in \mathbb{K}$. Then

$$(S \circ T(u + u')) = S(T(u + u')) = S(T(u)) + S(T(u')) = (S \circ T)(u) + (S \circ T)(u').$$

Thus $S \circ T$ is linear.

**Theorem 1.32.** Suppose that $T : U \to V$ and $S : V \to W$ are linear maps of $\mathbb{K}$-vector spaces. Then

1. If $S \circ T$ is surjective, then $S$ is surjective.
2. If $S \circ T$ is injective, then $T$ is injective.
3. If $S$ and $T$ are injective, then $S \circ T$ is injective.
4. If $S$ and $T$ are surjective, then $S \circ T$ is surjective.
5. If $S$ and $T$ are isomorphisms, then $S \circ T$ is an isomorphism.

**Proof.** Suppose $S \circ T$ is surjective and $W \in W$. Then there is some $u \in U$ such that $(S \circ T)(u) = w$. Let $v = T(u)$. Then $S(v) = w$. Thus $S$ is surjective.

Suppose that $S \circ T$ is injective and $T(u) = T(u')$. Then

$$(S \circ T)(u) = S(T(u)) = S(T(u')) = (S \circ T)(u').$$

It follows that $u = u'$, so $T$ is injective.

Suppose that $S$ and $T$ are injective and that $(S \circ T)(u) = (S \circ T)(w)$. Then $S(T(u)) = S(T(u'))$, and since $S$ is injective, this implies that $T(u) = T(u')$. But since $T$ is also injective, this implies that $u = u'$. It follows that $S \circ T$ is injective.

Suppose that $S$ and $T$ are surjective and that $w \in W$. Since $S$ is surjective, there is some $v \in V$ such that $S(v) = w$. Since $T$ is surjective, there is some $u \in U$ such that $T(u) = v$. But then we see that $(S \circ T)(u) = w$. It follows that $S \circ T$ is surjective.

Finally, if $S$ and $T$ are isomorphisms, then both $S$ and $T$ are injective, so that $S \circ T$ is injective, and both $S$ and $T$ are surjective, so that $S \circ T$ is surjective. From this we conclude that $S \circ T$ is an isomorphism.

**Theorem 1.33.** Let $B$ be a basis of $U$, $C$ be a basis of $V$ and $D$ be a basis of $W$. Suppose that $S : V \to W$ and $T : U \to V$ are linear transformations of $\mathbb{K}$-vector spaces. Then

$$(S \circ T)_{D,B} = S_{D,C}T_{C,B}.$$ 

Thus function composition is represented by matrix multiplication.
**Example 1.34.** Let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ be given by

$$T(x, y, z) = (3x - 2y + z, x - y + z, 3x - 5y + 4z, 6x - 3y - 4z)$$

and $S : \mathbb{R}^4 \rightarrow \mathbb{R}^5$ be given by

$$S(x, y, z, w) = (6x + 2y - 4z, 3x + 4y - 5z + 2w, 5x - 5y + w, -x - 3y + 4z - 7w, 4y - 7z - 6w).$$

Denote the standard bases of $\mathbb{R}^3$, $\mathbb{R}^4$ and $\mathbb{R}^5$ by $S$, $S'$ and $S''$ resp. Then

$$S_{S', S} = \begin{bmatrix} 6 & 2 & -4 & 0 \\ 3 & 4 & -5 & 2 \\ 5 & -5 & 0 & 1 \\ -1 & -3 & 4 & -7 \\ 0 & 4 & -7 & -6 \end{bmatrix}, \quad T_{S', S} = \begin{bmatrix} 3 & -2 & 1 \\ 1 & -1 & 1 \\ 3 & -5 & 4 \\ 6 & -3 & -4 \end{bmatrix},$$

so that $(T \circ S)_{S', S} = \begin{bmatrix} 16 & -8 & -4 \\ 10 & 9 & -21 \\ -36 & 6 & 40 \\ -53 & 49 & 0 \end{bmatrix}$ which is the product of the two matrices above.

**Proposition 1.35.** Let $V$ be a finite dimensional vector space and $B$ be a basis of $V$. Then the matrix of the identity $I_V$ with respect to the same input and output basis $B$ is the identity matrix $I$, in other words, $(I_V)_{B,B} = I$.

**Theorem 1.36.** Let $T : V \rightarrow W$ be an isomorphism and suppose that $V$ and $W$ are finite dimensional. Then $T^{-1} : W \rightarrow V$ is a linear transformation. If $A$ is the matrix of $T$ with respect to some input basis $B$ of $V$ and output basis $C$ of $W$, then $A = T_{C,B}$. Then the matrix of $T^{-1}$ with respect to the input basis $C$ and output basis $B$ is $A^{-1}$.

**Proof.** Let $T : V \rightarrow W$ be an isomorphism. Since $T$ is bijective, $T^{-1}$ exists. To show $T^{-1}$ is linear, suppose that $w_1, w_2 \in W$, and $c \in \mathbb{K}$. Then $w_1 = T(v_1)$ and $w_2 = T(v_2)$ for some $v_1, v_2 \in V$. Moreover, $T^{-1}(w_1) = v_1$ and $T^{-1}(w_2) = v_2$, so we obtain

$$T^{-1}(w_1 + w_2) = T^{-1}(v_1) + T^{-1}(v_2) = T^{-1}(T(v_1) + T(v_2)) = v_1 + v_2 = T^{-1}(w_1) + T^{-1}(w_2).$$

That $T_{B,C}^{-1} = A^{-1}$ follows from the fact that $T^{-1}T = I_V$, so

$$I = (I_V)_{B,B} = (T^{-1})_{B,C}T_{C,B} = (T^{-1})_{B,C}A,$$

from which it follows immediately that $(T^{-1})_{B,C} = A^{-1}$. \qed
Let $V$ be a finite dimensional vector space with two bases $B$ and $B'$. The transition matrix $P$ from the basis $B$ to the basis $B'$ is the matrix of the identity with respect to the input basis $B$ and output basis $B'$. In other words, $P = I_{B',B}$.

**Proof.** By definition, $P$ satisfies $[v]_{B'} = P[v]_B$ for all $v \in V$. But
\[ I_{B',B}[v]_B = [I(v)]_{B'} = [v]_{B'}, \]
for all $v \in V$. It follows that the two matrices are defined by the same condition. $\square$

**Theorem 1.38.** Suppose that $T_{C,B}$ is the matrix of a linear transformation $T : V \to W$, with respect to the input basis $B$ of $V$, and output basis $C$ of $W$. If $B'$ is another basis of $V$ and $C'$ is another basis of $W$, then the matrix $T_{C',B'}$ of $T$ with respect to the input basis $B'$ and output basis $C'$ is given by
\[ T_{C',B'} = I_{C',C}T_{C,B}I_{B,B'}. \]

**Proof.** Using the properties of matrices of compositions of linear maps, we compute
\[ I_{C',C}T_{C,B}I_{B,B'} = (I_W \circ T)_{C',B'}I_{B,B'} = T_{C',B'}I_{B,B'} = (T \circ I_V)_{C',B'} = T_{C',B'}. \]

The theorem above allows us to compute the matrix of a linear transformation with respect to any bases of the input and output spaces as long as you can write down the matrix with respect to some input and output bases, since the process of computing transition matrices between bases is straightforward.

**Definition 1.39.** Let $A$ and $B$ be $n \times n$ matrices. Then we say that $B$ is similar to $A$ if there is an invertible matrix $P$ such that $B = P^{-1}AP$.

Note that $P$ must also be an $n \times n$ matrix.

**Theorem 1.40.** Similarity is an equivalence relation. In other words
\begin{enumerate}
  \item $A$ is similar to $A$ (The reflexive property)
  \item If $A$ is similar to $B$, then $B$ is similar to $A$. (Symmetry)
  \item If $A$ is similar to $B$ and $B$ is similar to $C$, then $A$ is similar to $C$. (Transitivity)
\end{enumerate}

**Proof.** Since $A = I^{-1}AI$, $A$ is similar to $A$, so the reflexive property holds. Suppose that $A$ is similar to $B$. Then there is a matrix $P$ such that $A = P^{-1}BP$. But then $B = PAP^{-1} = (P^{-1})^{-1}AP^{-1}$, so $B$ is similar to $A$. Thus symmetry holds. Finally, suppose that $A$ is similar to $B$ and $B$ is similar to $C$. Then there are matrices $P$ and $Q$ such that $A = P^{-1}BP$ and $B = Q^{-1}CQ$. Then
\[ A = P^{-1}BP = P^{-1}(Q^{-1}CQ)P = (QP)^{-1}C(QP), \]
so $A$ is similar to $C$ and transitivity holds. $\square$

**Theorem 1.41.** If $A$ is similar to $B$, then $\det(A) = \det(B)$.

**Proof.** Exercise. $\square$

**Theorem 1.42.** Let $V$ be a finite dimensional vector space and $T : V \to V$ be a linear operator. If $A$ is the matrix of $T$ with respect to and input and output basis $B$, and $C$ is the matrix of $T$ with respect to another input and output basis $B'$, then $\det(A) = \det(B)$. Thus if we define the determinant of $T$ by $\det(T) = \det(A)$, this is independent on the choice of basis, so depends only on $T$. 

Proof. Let $A = T_{B,B}$, $B = T_{B',B'}$, and $P = I_{B',B}$. Then $I_{B,B'} = P^{-1}$ so $A = P^{-1}BP$ and thus $A$ and $B$ are similar matrices. It follows that $\det(A) = \det(B)$. This implies that the determinant of a matrix representing $T$ with respect to the same input and output matrix does not depend on the choice of basis. Thus the determinant of $T$ is well defined. \hfill \Box

2. Eigenvalues and Eigenvectors

Definition 2.1. Let $T : V \to V$ be a linear operator on a vector space over $\mathbb{K}$. Then a nonzero element $v \in V$ is said to be an eigenvector corresponding to the eigenvalue $\lambda \in \mathbb{K}$ provided that $T(v) = \lambda v$. An element $\lambda \in \mathbb{K}$ is called an eigenvalue of $T$ if there is an eigenvector corresponding to it.

Definition 2.2. Let $A$ be an $n \times n$ matrix with coefficients in $\mathbb{K}$. Then an eigenvector $v$ of $A$ corresponding to the eigenvalue $\lambda \in \mathbb{K}$, is a nonzero $n \times 1$ column vector $v$ such that $Ax = \lambda x$. An element $\lambda \in \mathbb{K}$ is called an eigenvalue of $A$ provided that there is an eigenvector corresponding to it.

The two definitions of eigenvectors and eigenvalues are connected as follows.

Theorem 2.3. Let $T : V \to V$ be a linear operator on a finite dimensional vector space and $A$ be the matrix representing $T$ in terms of some basis $B$; i.e., $A = T_{B,B}$. Then a vector $v$ is an eigenvalue of $T$ with respect to the eigenvalue $\lambda$ precisely when $[v]_B$ is an eigenvector of $A$ with respect to $\lambda$.

Proof. Suppose that $v$ is an eigenvector of $T$ with respect to the eigenvalue $\lambda$. Then $T(v) = \lambda v$, so that

$$A[v]_B = T_{B,B}[v]_B = [T(v)]_B = [\lambda v]_B = \lambda [v]_B,$$

so $x = [v]_B$ is an eigenvector of $A$ with respect to the eigenvalue $\lambda$, since $x \neq 0$ because $v \neq 0$. On the other hand, suppose $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, and $B = (v_1, \ldots, v_n)$. If $x$ is an eigenvector of $A$ corresponding to $\lambda$, and we let $v = x_1 v_1 + \cdots + x_n v_n$, then $x = [v]_B$, so that

$$[T(v)]_B = A[v]_B = Ax = \lambda x = \lambda [v]_B = [\lambda v]_B,$$

which shows that $T(v) = \lambda v$. \hfill \Box

As a consequence of this theorem, we see that the eigenvectors and eigenvalues of a linear transformation correspond to the eigenvectors and eigenvalues of any matrix which represents $T$ in terms of some basis. This observation leads to the following corollary.

Corollary 2.4. Let $A \sim B$. Then $A$ and $B$ have the same eigenvalues. Moreover, if $B = B^{-1}A$ and $x$ is an eigenvector of $A$ corresponding to the eigenvalue $\lambda$, then $y = P^{-1}x$ is an eigenvector of $B$ corresponding to $\lambda$.

Proof. Let $Ax = \lambda x$ and $B = P^{-1}AP$. Then if $y = P^{-1}x$, we compute

$$By = (P^{-1}AP)^{-1}x = P^{-1}Ax = P^{-1}\lambda x = \lambda P^{-1}x = \lambda y.$$ 

Since $x \neq 0$, we also have $y \neq 0$, which means that $y$ is an eigenvector of $B$ corresponding to $\lambda$. Thus every eigenvalue of $A$ is an eigenvalue of $B$. Since similarity is an equivalence relation, it follows that every eigenvalue of $B$ is also an eigenvalue of $A$. Thus $A$ and $B$ have the same eigenvalues. \hfill \Box
As a consequence of the results above, studying eigenvectors and eigenvalues of linear operators on finite dimensional vector spaces can be reduced to the study of eigenvalues and eigenvectors of square matrices. Thus, if we analyze the matrix case, there are corresponding results for linear operators.

**Theorem 2.5.** Let $A$ be an $n \times n$ matrix and $B = \lambda I - A$. Then a nonzero vector $x$ is an eigenvector of $A$ precisely when $Bx = 0$. Thus $\lambda$ is an eigenvalue of $A$ precisely when $\det(\lambda I - A) = 0$.

**Proof.** Clearly $Bx = 0$ precisely when $Ax = \lambda x$. Thus a nonzero vector $x$ is an eigenvector of $A$ precisely when $x$ is in the nullspace of $B$. The nullspace of $B$ is nontrivial precisely when $\det B$ is zero, thus the eigenvalues of $A$ are precisely those $\lambda$ for which $\det(\lambda I - A) = 0$.

The consequences of this theorem are quite profound. The first one is that there are only a finite number of eigenvalues for a matrix.

**Theorem 2.6.** Let $A$ be an $n \times n$ matrix and $p_A(\lambda) = \det(\lambda I - A)$ be the characteristic polynomial of $A$, in the variable $\lambda$. Then $p_A$ is a polynomial of degree $n$, so has at most $n$ roots. As a consequence, an $n \times n$ matrix $A$ has at most $n$ eigenvalues.

**Example 2.7.** Let $A = \begin{bmatrix} -11 & -14 \\ 7 & 10 \end{bmatrix}$, with real coefficients. The characteristic polynomial $p_A(\lambda)$ is given by

$$p_A(\lambda) = \det(\lambda I - A) = \begin{vmatrix} -11 & -14 \\ 7 & 10 \end{vmatrix} = \lambda^2 + \lambda - 12 = (\lambda - 3)(\lambda + 4).$$

Thus $\lambda = 3$ and $\lambda = -4$ are the eigenvalues of $A$. To find corresponding eigenvectors, we first consider the eigenvalue 3, and form the matrix $3I - A = \begin{bmatrix} 14 & 14 \\ -7 & -7 \end{bmatrix}$, which has Reduced Row Echelon form $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Thus, we see that $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, is a corresponding eigenvector. One can check directly that $Av_1 = 3v_1$. A similar calculation reveals that $v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ is an eigenvector corresponding to the eigenvalue $-4$. In this, example, we see that $\langle v_1, v_2 \rangle$ is a basis of $\mathbb{R}^2$ consisting of eigenvectors of $A$.

One of the advantages of working over the complex numbers is that every polynomial factors completely into linear factors, so that in particular, there is always an eigenvalue for every $n \times n$ matrix. For the Real numbers, the situation is different.

**Example 2.8.** Let $A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$. This matrix represents a rotation of the plane by the angle $\theta$. Since the column vectors of $A$ are an orthonormal basis of $\mathbb{R}^2$, $A$ is an orthogonal matrix, and it is easily calculated that $\det(A) = 1$. In fact, these two conditions are necessary and sufficient for a $2 \times 2$ matrix to represent a rotation. From a geometrical perspective, it is not hard to realize that the only rotations which could have eigenvectors are the rotation by angle $\pi$ and the trivial rotation (by angle 0). Let us verify this.
\[
\det (\lambda I - A) = \begin{vmatrix}
\lambda - \cos \theta & \sin \theta \\
-\sin \theta & \lambda - \cos \theta
\end{vmatrix} = \lambda^2 - 2\lambda \cos(\theta) + 1.
\]

The roots of \( p_A(\lambda) = \lambda^2 - 2\lambda \cos(\theta) + 1 \) are

\[
\lambda = \cos(\theta) \pm \sqrt{\cos(\theta)^2 - 1}.
\]

These numbers are nonreal unless \( \cos(\theta) = \pm 1 \), which occurs at \( \theta = 0, \pi \).

This shows that for most rotation matrices, there are no real eigenvalues, but there are two distinct complex eigenvalues. Thus, the existence of eigenvectors and eigenvalues of this matrix depend on whether we are working over the field \( \mathbb{C} \) or the field \( \mathbb{R} \). This type of phenomenon occurs regularly, and is one of the reasons that in many cases, mathematicians prefer to work over the field \( \mathbb{C} \), because in this case, the problem is simpler.

**Definition 2.9.** Let \( \lambda \) be an eigenvalue for a linear operator \( T : V \to V \). Then the eigenspace \( V_\lambda \) is the set of all \( v \in V \) such that \( T(v) = \lambda v \).

For an \( n \times n \) matrix \( A \), we define the eigenspace corresponding to the eigenvalue \( \lambda \) to be the set of all \( n \times 1 \) vectors \( x \) such that \( Ax = 0 \).

Note that the eigenspace for \( \lambda \) consists of all the eigenvectors corresponding to the eigenvalue \( \lambda \) plus the zero vector, which is not an eigenvector because we exclude the zero vector in the definition of and eigenvector.

**Proposition 2.10.** The eigenspace \( V_\lambda \) corresponding to an eigenvalue for a linear operator \( T : V \to V \) is a subspace of \( V \). Similarly, the eigenspace corresponding to an eigenvalue \( \lambda \) of an \( n \times n \) matrix \( A \) is a subspace of \( M_{n,1} \).

*Proof.* Exercise.

**Definition 2.11.** An eigenbasis of a linear operator \( T : V \to V \) is a basis \( B \) of \( V \) consisting of eigenvectors of \( T \). Similarly, an eigenbasis of an \( n \times n \) matrix \( A \) is a basis of \( M_{n,1} \) consisting of eigenvectors of \( A \).

We often identify the space \( M_{n,1} \) with \( \mathbb{R}^n \), which is possible to do since both spaces have standard bases, allowing us to give a canonical isomorphism between the two spaces. Using this identification, we often consider eigenvectors of \( A \) to be vectors in \( \mathbb{R}^n \), and we will sometimes do this.

**Proposition 2.12.** Let \( A \) be an upper triangular, lower triangular, or diagonal matrix. Then the diagonal entries of \( A \) are the eigenvalues of \( A \).

*Proof.* We already know that the determinant of a triangular matrix is the product of the diagonal entries of the matrix. Suppose that the diagonal entries of \( A \) are \( \lambda_1, \ldots, \lambda_n \). Then noting that \( \lambda I - A \) is a triangular matrix, with diagonal entries \( \lambda - \lambda_1, \ldots, \lambda - \lambda_n \), we note that the determinant \( p_A(\lambda) \) of this matrix is

\[
p_A(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n),
\]

which has roots \( \lambda_1, \ldots, \lambda_n \). For each of these values \( \lambda_i \) of \( \lambda \), the determinant of \( \lambda_i I - A \) vanishes, which implies that its nullspace is nontrivial, so there is a nonzero element \( x \) in the nullspace of the matrix \( \lambda_i I - A \), which is an eigenvector of \( A \) corresponding to the eigenvalue \( \lambda_i \). In particular, the diagonal entries are the eigenvalues of \( A \).

**Definition 2.13.** An \( n \times n \) matrix \( A \) is called diagonalizable iff it is similar to a diagonal matrix; i.e., there is an invertible matrix \( P \) such that \( D = P^{-1}AP \) is a diagonal matrix.
Theorem 2.14. For an \( n \times n \) matrix, the following are equivalent.

(1) \( A \) is diagonalizable.

(2) There is a basis of \( \mathbb{R}^n \) consisting of eigenvectors of \( A \).

Proof. Suppose that \( A \) is diagonalizable and \( P \) is such that \( D = P^{-1}AP \) is a diagonal matrix. Then \( PD = AP \). Let \( P = [v_1 \cdots v_n] \). Then \( AP = [Av_1 | \cdots | Av_n] \).

Also, we compute \( PD = [\lambda_1v_1 | \cdots | \lambda_nv_n] \), where \( D = \text{Diag}(\lambda_1, \ldots, \lambda_n) \). It follows that \( Av_i = \lambda_i v_i \) for \( i = 1, \ldots, n \), so \( \{v_1, \ldots, v_n\} \) is a basis of \( \mathbb{R}^n \) consisting of eigenvectors of \( A \).

On the other hand, let \( \{v_1, \ldots, v_n\} \) be a basis of \( \mathbb{R}^n \) consisting of eigenvectors of \( A \) and let \( P = [v_1 | \cdots | v_n] \). Let \( \lambda_i \) be the eigenvalue corresponding to \( v_i \), so that \( Av_i = \lambda_i v_i \). Let \( D = \text{Diag}(\lambda_1, \ldots, \lambda_n) \). Now \( AP = [Av_1 | \cdots | Av_n] = [v_1 | \cdots | v_n]D = PD \), so it follows that \( D = P^{-1}AP \). Thus \( A \) is diagonalizable. \( \square \)

It follows from the proof above that if we choose any eigenbasis of \( A \), the matrix \( P \) constructed from this eigenbasis satisfies the property that the matrix \( D = P^{-1}AP \) is diagonal. In particular, since we could have chosen to order the elements of the eigenbasis in any order, the order of the eigenvalues in such a diagonalization \( D \) of \( A \) can be arbitrary.

Definition 2.15. Let \( A \) be an \( n \times n \) matrix. Then the exponential \( \exp(A) \) of \( A \) is the matrix

\[
\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k.
\]

The above definition has one problem, which is how do we interpret the infinite sum above. We will not prove that the sum converges in this class, as the techniques for its proof are more suited to an analysis class.

Theorem 2.16. Let \( D = \text{Diag}(\lambda_1, \ldots, \lambda_n) \) be a diagonal matrix. Then

\[
\exp(D) = \text{Diag}(\exp(\lambda_1), \ldots, \exp(\lambda_n)).
\]

As a consequence, if \( A \) is a diagonalizable matrix, and \( D = P^{-1}AP \), where \( D = \text{Diag}(\lambda_1, \ldots, \lambda_n) \) is diagonal, then we can compute

\[
\exp(A) = P \text{Diag}(\exp(\lambda_1), \ldots, \exp(\lambda_n))P^{-1}.
\]

Proof. Exercise. \( \square \)

As a consequence of this theorem, we can compute the exponential of a diagonalizable matrix by matrix explicitly, in terms of the exponentials of the eigenvalues and matrix multiplication.

Example 2.17. We apply the exponential of a matrix above to solve systems of differential equations. A system of linear differential equations with constant coefficients is a set of \( n \) differential equations in \( n \) variables \( x_i \) of the form

\[
\begin{align*}
\dot{x}_1 &= a_{11}x_1 + \cdots + a_{1n}x_n \\
&\vdots \\
\dot{x}_n &= a_{n1}x_1 + \cdots + a_{nn}x_n
\end{align*}
\]
where the dot stands for the derivative with respect to time. This system of differential equations can be written compactly in matrix form

\[ \dot{x} = Ax, \]

where \( A = (a_{ij}) \) and \( x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \). The general solution to this system of differential equations is given by \( x = \exp(tA)X_0 \), where \( X_0 \) is an arbitrary constant vector.

When \( A \) is diagonalizable, it is easy to give the solution explicitly, because we can explicitly compute \( \exp(tA) \).

### 3. Complex Inner Product Spaces

**Definition 3.1.** Let \( V \) be a complex vector space. Then a Hermitian inner product on \( V \) is a map \( \langle \cdot, \cdot \rangle : V \times V \to \mathbb{C} \) satisfying

1. \( \langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle. \) (Additivity)
2. \( \langle \lambda v, w \rangle = \lambda \langle v, w \rangle. \) (Homogeneity)
3. \( \langle w, v \rangle = \overline{\langle v, w \rangle}. \) (Conjugate Symmetry)
4. \( \langle v, v \rangle \geq 0 \) and \( \langle v, v \rangle = 0 \) iff \( v = 0 \) (Positive definiteness)

Physicists make a slight variation on this definition, by replacing the second condition with the rule \( \langle v, \lambda w \rangle = \lambda \langle v, w \rangle. \) Their definition is equivalent to the one that mathematicians use, and one can transform a Hermitian inner product in mathematics formulation to the physics formulation by replacing the inner product by its conjugate.

**Proposition 3.2.** Let \( V \) be a complex inner product space. Then

1. \( \langle 0, v \rangle = 0 \) for all \( v \in V \).
2. \( \langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle. \)
3. \( \langle v, \lambda w \rangle = \overline{\lambda} \langle v, w \rangle. \)

**Proof.** To see the first statement, we use additivity to obtain that

\[ \langle 0, v \rangle = \langle 0 + 0, v \rangle = \langle 0, v \rangle + \langle 0, v \rangle. \]

Since \( \langle 0, v \rangle \) is a complex number, this forces \( \langle 0, v \rangle = 0 \).

To see the second assertion, we compute

\[ \langle v, w + w' \rangle = \overline{\langle w + w', v \rangle} = \overline{\langle w, v \rangle + \langle w', v \rangle} = \overline{\langle w, v \rangle} + \overline{\langle w', v \rangle} = \langle v, w \rangle + \langle v, w' \rangle. \]

To see the last assertion, note that

\[ \langle v, \lambda w \rangle = \langle \lambda w, v \rangle = \overline{\lambda} \langle w, v \rangle = \overline{\lambda} \langle w, v \rangle = \overline{\lambda} \langle v, w \rangle. \]


**Definition 3.3.** If \( V \) is a complex inner product space and \( v \in V \), then the norm of \( v \), denoted by \( \|v\| \), is given by \( \|v\| = \sqrt{\langle v, v \rangle} \).

**Definition 3.4.** Let \( V \) be a finite dimensional complex inner product space. Then a basis \( \langle e_1, \ldots, e_n \rangle \) of \( V \) is said to be an orthonormal basis of \( V \) provided that

\[ \langle e_i, e_j \rangle = \delta_{ij}, \]

where \( \delta_{ij} \) is the Kronecker Delta symbol.
Theorem 3.5. Let \( \langle e_1, \ldots, e_n \rangle \) be an orthonormal basis of a finite dimensional complex inner product space \( V \), and \( v \in V \). Then
\[
v = \sum_{i=1}^{n} \langle v, e_i \rangle e_i.
\]

Theorem 3.6 (Existence of an orthonormal basis). Let \( V = \langle v_1, \ldots, v_n \rangle \) be a finite dimensional complex inner product space, and consider the subspaces \( V_k = \langle v_1, \ldots, v_k \rangle \) for \( k = 1, \ldots, n \). Then there is a basis \( \langle e_1, \ldots, e_n \rangle \) of \( V \) such that \( \langle e_1, \ldots, e_k \rangle \) is an orthonormal basis of \( V_k \) for \( k = 1, \ldots, n \).

Proof. Let \( e_1 = \frac{1}{\|v_1\|} v_1 \). Then \( \|e_1\| = 1 \) and \( \langle e_1 \rangle \) must be a basis of \( V_1 \), since \( \langle v_1 \rangle \) is a basis of \( V_1 \). Now, suppose we have been able to construct an orthonormal basis \( \langle e_1, \ldots, e_k \rangle \) of \( V_k \) so that \( \langle e_1, \ldots, e_\ell \rangle \) is a basis of \( V_\ell \) for all \( 1 \leq \ell \leq k \), and that \( k < n \). Let
\[
u = v_{k+1} - \sum_{i=1}^{k} \langle v_{k+1}, e_i \rangle e_i.
\]
Then
\[
\langle u, e_j \rangle = \langle v_{k+1}, e_j \rangle - \sum_{i=1}^{k} \langle v_{k+1}, e_i \rangle e_i, e_j = 0.
\]
Set \( e_{k+1} = \frac{1}{\|u\|} u \). It is easy to check that \( \langle e_1, \ldots, e_{k+1} \rangle \) is an orthonormal basis of \( V_{k+1} \).

The process of constructing an orthonormal basis of a complex inner product space is the same as the Gram-Schmidt process for a real inner product space.

Theorem 3.7. Let \( T : V \to V \) be a linear operator on a finite dimensional complex inner product space. Then there is a unique linear operator \( T^* : V \to V \), called the adjoint of \( T \), which satisfies
\[
\langle T^*(v), w \rangle = \langle v, T(w) \rangle
\]
for all \( v, w \in V \).

Proof. Let \( \langle e_1, \ldots, e_n \rangle \) be an orthonormal basis of \( V \). Define
\[
T^*(e_i) = \sum_{j=1}^{n} \langle e_i, T(e_j) \rangle e_j.
\]
Since we can define a linear map uniquely by specifying its action on a basis, this defines a linear map \( T^* \). We show this linear map satisfies the required condition. Let \( v = \sum_{i=1}^{n} a_i e_i \) and \( w = \sum_{k=1}^{n} b_k e_k \). Then
\[
T^*(v) = T^*\left( \sum_{i=1}^{n} a_i e_i \right) = \sum_{i=1}^{n} a_i T^*(e_i) = \sum_{i,j=1}^{n} a_i \langle e_i, T(e_j) \rangle e_j,
\]
so that
\[
\langle T^*(v), w \rangle = \sum_{i,j,k=1}^{n} a_i \langle e_i, T(e_j) \rangle \langle e_j, b_k e_k \rangle = \sum_{i,j=1}^{n} a_i b_j \langle e_i, T(e_j) \rangle = \langle v, T(w) \rangle.
\]
Thus, the linear operator $T^*$ satisfies the required conditions. To see such an operator is unique, suppose that $S$ is another linear operator satisfying $\langle S(v), w \rangle = \langle v, T(w) \rangle$ for all $v, w \in V$. Then $(T^*(v), w) = \langle S(v), w \rangle$, so $\langle (S - T^*)(v), w \rangle = 0$ for all $w$. In particular, let $w = (S - T^*)(v)$, and then we obtain $\langle (S - T^*)(v), (S - T^*)(v) \rangle$, which forces $(S - T^*)(v) = 0$ for all $v \in V$. From this, we can conclude that $S = T^*$. \hfill \Box

**Example 3.8.** Let $V = \mathbb{C}^n$, equipped with the standard Hermitian inner product

$$\langle z, w \rangle = \sum_{i=1}^{n} z_i \overline{w}_i,$$

If we consider an element of $\mathbb{C}^n$ to be an $n \times 1$ matrix $z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$, then we can express the inner product as

$$\langle z, w \rangle = \overline{w}^T v.$$

If a linear operator $T$ on $V$ is given by $T(z) = Az$, then we claim that the matrix of $T^*$ is just $A^T$, which is called the conjugate transpose of $A$. To see this, suppose that $T^*(z) = Bz$. Then we have

$$\langle z, Aw \rangle = \langle z, T(w) \rangle = \langle T^*(z), w \rangle = \overline{w}^T Bz = (B^T w)^T z = \langle z, B^T w \rangle,$$

which forces $A = B^T$, and thus $B = A^T$.

**Definition 3.9.** A linear operator $T$ on a complex inner product space is called **self adjoint** (or Hermitian) if $T = T^*$. Similarly, a complex matrix $A$ is **Hermitian** if $A = A^T$.

When $V$ is infinite dimensional, the term self adjoint usually is only applied to linear operators that satisfy an additional topological condition. We will not discuss this issue here, but it is important in physics applications, where the operators are on an infinite dimensional Hilbert space.

**Theorem 3.10** (Spectral Theorem for Hermitian Operators on a Finite Dimensional Space). Let $T$ be a self adjoint operator on a finite dimensional complex inner product space $V$. Then

1. Every eigenvalue of $T$ is real.
2. If $W = \langle v \rangle$, where $v$ is an eigenvector of $T$, then for any $w \in W^\perp$, $T(v) \in W^\perp$. Thus $T$ induces a map $W^\perp \rightarrow W^\perp$, and this induced map is self adjoint on the induced inner product space structure of $W^\perp$.
3. There is an orthonormal basis of $V$ consisting of eigenvectors of $T$.

**Proof.** Let $v$ be an eigenvector corresponding to the eigenvalue $\lambda$. If $\lambda = 0$, then $\lambda$ is evidently real, so assume that $\lambda \neq 0$. Then

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle T(v), v \rangle = \langle v, T(v) \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle.$$ 

Since $\langle v, v \rangle \neq 0$, it follows that $\overline{\lambda} = \lambda$, which means that $\lambda$ is real. Now let $W = \langle v \rangle$ and suppose $w \in W^\perp$. Then

$$\langle v, T(w) \rangle = \langle T(v), w \rangle = \langle \lambda v, w \rangle = \lambda \langle v, w \rangle = 0,$$
so it follows that $T(w) \in W^\perp$. Let $S : W^\perp \to W^\perp$ be given by $S(w) = T(w)$ for $w \in W^\perp$. Then we have for any $w, w' \in W^\perp$

$$\langle S^*(w), w' \rangle = \langle w, S(w') \rangle = \langle w, T(w') \rangle = \langle T(w), w' \rangle = \langle S(w), w' \rangle.$$  

It follows that $S = S^*$, so $S$ is self adjoint.

If $\dim(V) = 1$, then let $V = \langle v \rangle$. Then $T(v) = \lambda v$ for some $\lambda$, so $v$ is an eigenvector of $T$, and we can normalize $v$ to obtain an orthonormal basis of $V$. Suppose that the theorem has been shown whenever $\dim(V) = k$ and $\dim(V) = k + 1$. Now, $T$ can be represented by a complex matrix $A$ with respect to some basis of $V$, and every complex square matrix must have an eigenvalue, since its characteristic polynomial $p_A(\lambda)$ has to have at least one root. But this means that there must be an eigenvector $v$ of $T$. Let $W = \langle v \rangle$. Let $S : W^\perp \to W^\perp$ be the induced map, so that by condition 2, we know that $S$ is self adjoint. Since $\dim(W^\perp) = k$, we can find a orthonormal basis of $W^\perp$ consisting of eigenvectors of $S$. However, since $S = T$ on $W^\perp$, these vectors are also eigenvectors of $T$. Normalizing $v$, we get an eigenvector of $T$ which is a multiple of $v$. The combined set consisting of the orthonormal basis of $W^\perp$ and the normalize basis of $W$ gives the required orthonormal basis of $V$ consisting of eigenvectors of $T$. 

The terminology “Spectral Theorem” first appeared in a paper of Hilbert, but later, physicists related the spectral lines of atoms to eigenvalues of certain operators, so there turned out to be a serendipitous coincidence of terminology. The set of eigenvalues of a Hermitian operator on a finite dimensional space is called the spectrum of the operator. For infinite dimensional operators, the situation is more complicated, and we will not discuss this further here.

**Theorem 3.11.** Let $A$ be a Hermitian matrix. Then

1. Every eigenvalue of $A$ is real.
2. $A$ can be orthogonally diagonalized. That is, there is an orthogonal matrix $P$ such that $D = P^{-1}AP$ is diagonal.

**Proof.** Define $T$ on $\mathbb{C}^n$ by $T(z) = Az$. Then $T$ is a self adjoint operator, so there is an orthonormal basis $\langle u_1, \ldots, u_n \rangle$ of $\mathbb{C}^n$ consisting of eigenvectors of $T$. Their coordinate vectors are eigenvectors of $A$. Let $P$ be the matrix obtained from these coordinate vectors, that is $P = [u_1 | \ldots | u_n]$. Then $P$ is an orthogonal matrix. Suppose that $Au_i = \lambda_i u_i$. Let $D = \text{Diag}(\lambda_1, \ldots, \lambda_n)$. Then we have

$$PD = [\lambda_1 u_1 | \ldots | \lambda_n u_n] = [Au_1 | \ldots | Au_n] = AP.$$  

Thus $D = P^{-1}AP$. The fact that any eigenvalue of $A$ is real is immediate from the observation that $T$ and $A$ have the same eigenvalues, and $T$ is self adjoint, so all its eigenvalues are real. 

**Department of Mathematics, University of Wisconsin-Eau Claire, Eau Claire, WI 54729 USA**

E-mail address: penkavmr@uwec.edu