# LINEAR ALGEBRA

SIDDHANT CHAUDHARY

Abstract. In this document, I discuss topics in linear algebra. I have also included solutions to some selected exercises from the book *Linear Algebra (Third Edition)* by *Serge Lang*.

## I, §1. Exercises

**1.** Let $V$ be a vector space over a field $K$. Let $c \in K$. We show that $c\mathbf{0} = \mathbf{0}$. The proof is as follows:

$$\begin{aligned} c\mathbf{0} &= c(\mathbf{0} + \mathbf{0}) \\ &= c\mathbf{0} + c\mathbf{0} \end{aligned}$$

and by adding the inverse of $c\mathbf{0}$ on both sides, we may obtain the required equality.

**2.** Suppose $c \in K$ such that $c \neq 0$, and let $\mathbf{v} \in V$. Also, suppose $c\mathbf{v} = \mathbf{0}$. Then, by multiplying $c^{-1}$ on both sides ($K$ is a field), we may obtain that $\mathbf{v} = \mathbf{0}$.

**6.** Before solving this problem, we prove a theorem: Let $V$ be a vector space over some field $K$, and let $U$, $W$ be $K$-subspaces of $V$. Then, $U \cap W$ is a $K$-subspace of $V$.

*Proof*: It is clear to see that $\mathbf{0} \in U \cap W$. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $U \cap W$, so is $\mathbf{u} + \mathbf{v}$. If $c \in K$, and if $\mathbf{u}$ is a vector in $U \cap W$, so is $c\mathbf{u}$. Hence, $U \cap W$ is a $K$-subspace of $V$.

This problem now directly follows from this theorem.

## A Few Points About Dimension

Here, I won't prove any theorem about the dimension of a vector space. But, I just want to plot a chain of reasoning which leads to beautiful statements about the dimension. If $V$ is a vector space over a field $K$, and suppose it is finite dimensional, we denote its dimension by $\dim_K V$

(1) Suppose $\{v_1, v_2, v_3, ...v_n\}$ is a basis for $V$. Then, any subset of $V$ the size of which is greater than $n$ must be linearly dependent. The proof of this fact has the following sketch: assume that the subset is linearly independent, and show that a smaller subset of that subset generates $V$, thereby getting a contradiction.
(2) This means that any two bases of $V$ must have the same size. This is why the dimension is a well-defined number.
(3) Any maximal subset of linearly independent elements of $V$ constitutes a basis.

## Extension to Infinite Dimensional Spaces

Here, we'll try to extend the notion of a basis even to infinite dimensional spaces. First, let's start with a definition.

*Date*: August 2019.

**Definition:** Suppose $V$ is a vector space over a field $K$, and suppose $X \subseteq V$. We denote by $\langle X \rangle$ the smallest subspace of $V$ that contains $X$, and call it the *span* of $X$.

We will now show that $\langle X \rangle$ always exists.

**Theorem 0.1.** Given $X \subseteq V$ such that $X$ is non-empty, $\langle X \rangle$ exists.

*Proof:* Let $S$ be the set of all subspaces of $V$ that contain $X$. Clearly, $V \in S$, and thus $S$ is non-empty. Then, it is not hard to see that the set $M$ defined by

$$M = \bigcap_{W \in S} W$$

is a subspace and contains $X$. Also, if $W$ is a subspace that contains $X$, then $M \subseteq W$, and so the theorem is proved.

If $X$ is finite, $\langle X \rangle$ cooincides with the usual definition of span, which we prove in the next theorem.

**Theorem 0.2.** Let $X \subseteq V$ be a finite set, say $\{v_1, v_2, ..., v_n\}$. Then, $\langle X \rangle$ is the set of all linear combinations of elements of $X$.

*Proof:* Consider the set $Q = \{a_1v_1 + a_2v_2 + ... + a_nv_n : a_i \in K\}$. Clearly, $Q$ is a subspace that contains $X$. So, $\langle X \rangle \subseteq Q$. Now, by definition, note that $v_i \in \langle X \rangle$ for each $i$, and since $\langle X \rangle$ is a subspace, it follows that

$$a_1v_1 + a_2v_2 + ... + a_nv_n \in \langle X \rangle$$

for any constants $a_i \in K$. This means that $Q \subseteq \langle X \rangle$, and hence we have

$$Q = \langle X \rangle.$$

The next theorem is a beautiful way of thinking about what the span of an infinite set $X$ is.

**Theorem 0.3.** Suppose $X \subseteq V$, $X$ is non-empty, and let $P$ be the set of all *finite* subsets of $X$. Then,

$$\langle X \rangle = \bigcup_{Y \in P} \langle Y \rangle$$

*Proof:* Set $M = \bigcup_{Y \in P} \langle Y \rangle$. First, lets show that $M$ is a subspace. Let $x$ and $y$ be in $M$. Then, $x \in \langle Y_1 \rangle$ for some finite subset $Y_1$ of $X$, and similarly $y \in \langle Y_2 \rangle$ for some finite subset $Y_2$ of $X$. Let $Y_1 = \{v_1, v_2, ..., v_s\}$ and let $Y_2 = \{u_1, u_2, ..., u_t\}$. Then, by **Theorem 0.2**, $x$ is a linear combination of elements of $Y_1$, and $y$ is a linear combination of elements of $Y_2$. This means that $x + y$ is a linear combination of elements of $Y_1 \cup Y_2$. But, $Y_1 \cup Y_2$ is a finite subset of $X$, and hence $x + y \in \langle Y_1 \cup Y_2 \rangle$, which means that $x + y \in M$. Similarly, it may be proved that $cx \in M$, where $c \in K$. Now, it is clear that $M$ contains $X$. This means that $\langle X \rangle \subseteq M$. Also, if $x \in M$, then $x$ is a linear combination of finitely many elements of $X$. Since $\langle X \rangle$ is a subspace containing each of those elements, it follows that $x \in \langle X \rangle$, which means that $M \subseteq \langle X \rangle$. Hence, $M = \langle X \rangle$.

The above theorem means that the span of a set $X$ is the set of *all* linear combinations of *finitely* many elements of $X$. This is a great way of extending the meaning of span to infinite sets.

Observe that the definition of Linear Independence remains that same for a finite subset $X$ of $V$. For an infinite subset $X$, we say that $X$ is linearly independent if every finite subset of $X$ is linearly independent. The definition of a maximal linearly independent set remains the same. Now, we introduce another definition.

**Definition:** Let $V$ be a vector space, and let $X \subseteq V$ . We say that $X$ is a *minimal spanning/generating* set of $V$ if $\langle X \rangle = V$ and if $Y$ is a proper subset of $X$, then $\langle Y \rangle \neq V$.

Now, we give a theorem which generalises the notion of a basis to arbitrary vector spaces, finite or infinite dimensional.

**Theorem 0.4.** Let $V$ be a vector space, and let $X \subseteq V$. Then, the following statements are equivalent.

(1) $X$ is a maximal linearly independent set of $V$.
(2) $X$ is a minimal spanning set of $V$.
(3) $X$ is linearly independent and $\langle X \rangle = V$.
(4) Every $v \in V$ is uniquely representable as

$$v = a_1 v_1 + a_2 v_2 + ... + a_n v_n$$

where $a_i \in K$, and $v_1, v_2, ..., v_n$ are some elements of $X$.

*Proof:* We will show that (1) $\implies$ (2) $\implies$ (3) $\implies$ (4) $\implies$ (1), which will complete the proof.

First, lets show that (1) $\implies$ (2). Suppose $X$ is a maximal linearly independent set in $V$. Let $y \in V$. Then, if $y \in X$, then by definition, $y \in \langle X \rangle$. If $y \notin X$, then the set $X \cup \{y\}$ is linearly dependent. So, there exist some $v_1, v_2, ..., v_n$ in $X$ and $a_1, a_2, ..., a_n, a_{n+1}$ in $K$ not all zero such that

$$a_1 v_1 + a_2 v_2 + ... a_n v_n + a_{n+1} y = O$$

and we know that $a_{n+1} \neq 0$. Hence,

$$y = -\frac{a_1}{a_{n+1}} v_1 - \frac{a_2}{a_{n+1}} v_2 - ... - \frac{a_n}{a_{n+1}} v_n$$

and hence $\langle X \rangle = V$. Now, let $Y$ be a proper subset of $X$, and let $v \in X - Y$. Then, $v$ cannot be represented as a finite linear combination of elements of $Y$ (because $X$ is linearly independent), and hence $v \notin \langle Y \rangle$, and hence $\langle Y \rangle \neq V$. So, $X$ is a minimal spanning set of $V$.

Now lets prove that (2) $\implies$ (3). So, let $X$ be a minimal spanning set of $V$. Then, by definition, $\langle X \rangle = V$, and we only need to show that $X$ is linearly independent. Suppose, on the contrary, that $X$ is linearly dependent. Then, there exist some $v_1, v_2, ..., v_n$ in $X$ and $a_1, a_2, ..., a_n$ in $K$ not all zero such that

$$a_1 v_1 + a_2 v_2 + ... + a_n v_n = O$$

Without loss of generality, suppose $a_1 \neq 0$, and thus $v_1$ is a linear combination of $v_2, v_3, ..., v_n$. But, this means that the set $X - \{v_1\}$ spans $V$, which is not possible since $X$ is a minimal spanning set of $V$. Thus, our assumption is wrong, and hence $X$ must be linearly independent.

Lets show that (3) $\implies$ (4). Suppose $X$ is linearly independent and $\langle X \rangle = V$. Let $v$ be a non-zero element of $V$. It is clear that $v$ is a linear combination of finitely many elements of $X$. We need to prove that this representation is unique. So, let $a_1, a_2, ..., a_n, b_1, b_2, ..., b_m$ be in $K$, and let $v_1, v_2, ..., v_n, u_1, u_2, ..., u_m$ be in $X$ such that each $a_i$ and $b_i$ is non-zero, and

$$v = a_1 v_1 + a_2 v_2 + ... + a_n v_n = b_1 u_1 + b_2 u_2 + ... + b_m u_m$$

which means that

$$a_1 v_1 + a_2 v_2 + ... + a_n v_n - b_1 u_1 - b_2 u_2 - ... - b_m u_m = O$$

Now, we show that $u_1 = v_i$ for some $i$. If that is not the case, then it would imply that $b_1 = 0$, since $X$ is linearly independent, which is a contradiction. So, without loss of generality, suppose $u_1 = v_1$. We can continue this method and argue that $u_2 = v_2$, $u_3 = v_3$ and in fact $m = n$, and $u_n = v_n$. So, it follows that

$$(b_1 - a_1)v_1 + (b_2 - a_2)v_2 + ...(b_n - a_n)v_n = 0$$

and since $X$ is linearly independent this implies that $b_i = a_i$. Thus, each $v \in V$ has a unique linear representation.

Finally, let's prove that (4) $\implies$ (1). Suppose every non-zero $v$ can be uniquely written as

$$v = a_1v_1 + a_2v_2 + ... + a_nv_n$$

for some $a_1, a_2, ..., a_n$ in $K$ and some $v_1, v_2, ..., v_n$ in $X$. First, observe that $O \notin X$, because otherwise representations will not be unique. Now, we show that $X$ is linearly independent. If $X$ was linearly dependent, there would be some $v, v_1, v_2, ..., v_n$ in $X$ and $a_1, a_2, ..., a_n$ not all zero in $K$ such that

$$v = a_1v_1 + a_2v_2 + ... + a_nv_n$$

Note that $1v = v$ is also a representation of $v$. Thus, this shows that the representation of $v$ is not unique, and hence our assumption was wrong. So, $X$ must be linearly independent.

To show that $X$ is a maximal linearly independent set, suppose there is some $y \in V - X$, and consider the set $X \cup \{y\}$. If $y = O$, then the set if clearly linearly dependent. If $y \neq O$, then $y$ can be uniquely written as

$$y = a_1v_1 + a_2v_2 + ... + a_nv_n$$

for some $a_1, a_2, ..., a_n$ in $K$, and some $v_1, v_2, ..., v_n$ in $X$, and note that $v_i \neq y$. Also, since $y \neq O$, this means that one of $a_i$ is non-zero. And this gives us linear dependence for the set $X \cup \{Y\}$. So, $X$ is a maximal linearly independent set.

Next, we will show that every non-zero vector space has a basis. We will use what is known as Zorn's Lemma, which is equivalent to the Axiom of Choice. Before that, let us introduce some definitions.

Definition: Let $X$ be a partially ordered set. A subset $C \subseteq X$ is called a *chain* if given any $x \in C$ and $y \in C$, either $x \geq y$ or $y \geq x$. For a subset $Y$ of $X$, an element $x_0 \in X$ is called an *upper bound* of $Y$ is $y \leq x_0$ for all $y \in Y$. An element $x \in X$ is called a maximal element of $X$ if for any $y \in X$, $y \geq x$ implies that $y = x$.

The statement of Zorn's Lemma is as follows:

**Zorn's Lemma:** Let $X$ be a poset, such that every chain $C$ has an upper bound in $X$. Then, $X$ contains a maximal element.

Let's now prove that every non-zero vector space has a basis.

**Theorem 0.5.** Let $V$ be a vector space over some field $K$. Then, given a linearly independent subset $S$ of $V$, $S$ is contained in some maximal linearly independent subset of $V$.

*Proof:* Let $S$ be a linearly independent subset of $V$. Define

$$X := \{T \subseteq V : T \text{ is linearly independent, } S \subseteq T\}$$

Since $S \in X$, $X$ is non-empty. Let's define a partial order in $X$. For $T_1$ and $T_2$ in $X$, we say that $T_1 \leq T_2$ if $T_1 \subseteq T_2$. It is not hard to see why this is a partial order.

If we show that $X$ has a maximal element w.r.t this partial order, then that maximal element will be linearly dependent, and it will contain $S$. Also, this maximal element will also be a maximal linearly independent subset of $V$. So, let's show that $X$ has a maximal element w.r.t the partial order.

Let $C$ be a chain in $X$. Put

$$M = \bigcup_{T \in C} T$$

Let us show that $M$ is linearly independent. Suppose not. Then, there exist elements $v_1, v_2, ..., v_n$ in $M$ and scalars $a_1, a_2, ..., a_n$ not all zero such that

$$a_1 v_1 + a_2 v_2 + ... + a_n v_n = O$$

Also, $v_1 \in T_1$, $v_2 \in T_2$, ..., $v_n \in T_n$ for some $T_1, T_2, ..., T_n$ in $C$. Since $C$ is a chain, this means that there is some $1 \leq i \leq n$ such that $v_1, v_2, ..., v_n$ are in $T_i$. But, this is a contradiction because $T_i$ is a linearly independent subset of $V$. So, $M$ must be linearly independent. Also, observe that for any $T \in C$, $T \leq M$, which shows that $C$ has an upper bound. So, every chain in $X$ has an upper bound, and thus $X$ has a maximal element. This maximal element of $X$ is a maximal linearly independent subset of $V$ that contains $S$.

As a corollary to this theorem, we can immediately conclude that every non-zero vector space has a basis.

Next, let us show that every spanning set of a vector space contains a basis.

**Theorem 0.6.** Let $V$ be a vector space over a field $K$, such that $V \neq \{O\}$. Let $S$ be a spanning set of $V$. Then, $S$ contains a basis of $V$.

Proof: We will do a very similar proof as we did in **Theorem 0.5**.

If $S$ is linearly independent then we are done. So, assume that $S$ is not linearly independent. Define the poset

$$X := \{U \subseteq S : U \text{ is linearly independent}\}$$

It is clear that $X$ is non-empty. Again, the partial order in $X$ is defined as: if $T_1$ and $T_2$ are in $X$, we say that $T_1 \leq T_2$ if $T_1 \subseteq T_2$. We will show that $X$ contains a maximal element w.r.t this partial order. Then, it can be easily shown that the span of this maximal element is $V$, which will mean that this maximal element is a basis of $V$, and at the same time this maximal element is a subset of $S$, which will complete the proof of the theorem.

Let $C$ be a chain in $X$. Then, define

$$M = \bigcup_{T \in C} T$$

That $M \subseteq S$ is clear. We will show that $M$ is also linearly independent, which will give us an upper bound for this chain $C$. Suppose $M$ is linearly dependent. Then, there are $v_1, v_2, ..., v_n$ in $M$ and $a_1, a_2, .., a_n$ in $K$ not all zero such that

$$a_1 v_1 + a_2 v_2 + ... + a_n v_n = O$$

Also, $v_i \in T_i$, for each $1 \leq i \leq n$, where $T_i$ is in $C$. Since $C$ is a chain, this means that each $v_i$ is in $T_j$, for some $1 \leq j \leq n$. But, this already gives us a contradiction, because $T_j$ is a linearly independent set. Hence, $M$ must be linearly independent. So, $M \in X$, and this is the upper bound for $C$.

By Zorn's Lemma, this means that $X$ must contain a maximal element. Let this maximal element be $H$. Since $H$ is linearly independent, the set $S - H$ is non-empty (because $S$ is linearly dependent). So, take any $v \in S - H$, and consider the set

$$H \cup \{v\}$$

Since $H$ is a maximal element of $X$ w.r.t the partial order, it must be true that $H \cup \{v\}$ is linearly dependent. But, this means that $v$ is a linear combination of finitely many elements of $H$. So, this means that every element of $S$ is a linear combination of finitely many elements of $H$. Since the span of $S$ was $V$, this means that the span of $H$ is also $V$, thereby proving that $H$ is a basis of $V$.

Let's prove a stronger version of the previous theorem.

**Theorem 0.7.** Let $V$ be a vector space over a field $K$. Suppose $S_2$ is a spanning set of $V$, and $S_1$ is a linearly independent subset of $V$ such that $S_1 \subseteq S_2$. Then, there is a basis $\beta$ of $V$ such that $S_1 \subseteq \beta \subseteq S_2$.

*Proof:* If $S_2$ is linearly independent then there is nothing to prove. So, let's assume that $S_2$ is linearly dependent. Now, define the set

$$X := \{U \subseteq S_2 : S_1 \subseteq U,\ U \text{ is linearly independent}\}$$

Clearly, $S_1 \in X$, and so $X$ is non-empty. Again, our partial order on $X$ remains the same. We will show that $X$ has a maximal element w.r.t this partial order, and we will show that this maximal element is a basis of $V$.

Let $C$ be a chain in $X$. Define

$$M = \bigcup_{T \in C} T$$

Clearly $S_1 \subseteq M$, and $M \subseteq S_2$. By the same reasoning as we did in **Theorem 0.6**, $M$ is linearly independent. So, $M \in X$, and $M$ is an upper bound for this chain $C$.

By Zorn's Lemma, $X$ has a maximal element w.r.t this partial order. Let the maximal element be $H$. So, $H$ is a maximal element of $X$. By similar reasoning as we did in **Theorem 0.6**, it follows that $H$ is a basis of $V$. So, the theorem is proved.

## II, §3. Exercises

**18.** Let $A$ be a square matrix.

**(a):** Suppose $A^2 = O$. Now, observe that

$$(I - A)(I + A) = I^2 - A^2 = I$$

so that $(I - A)$ is invertible.

**(b):** We can extend this trick. Suppose $A^3 = O$. Then, observe that

$$(I - A)(I + A + A^2) = I^3 - A^3 = I$$

which again means that $(I - A)$ is invertible.

**(c):** In general, if $A^n = O$, then we have

$$(I - A)(I + A + A^2 + ... + A^{n-1}) = I^n - A^n = I$$

so that $(I - A)$ is invertible.

**(d):** Suppose $A^2 + 2A + I = O$. This means that $I = -2A - A^2$, which means that $I = A(-2I - A)$, which means that $A$ is invertible.

**(e):** Suppose that $A^3 - A + I = O$, and by using a similar argument, we can show that $A$ is invertible.

**27.** Let $A$ and $B$ be $n \times n$ matrices given by:

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & ... & b_{1n} \\ b_{21} & b_{22} & ... & b_{2n} \\ ... & ... & ... & ... \\ b_{n1} & b_{n2} & ... & b_{nn} \end{pmatrix}$$

Now, a diagonal element $[AB]_{ii}$ is given by

$$[AB]_{ii} = \sum_{j=1}^{n} a_{ij} b_{ji}$$

and a diagonal element $[BA]_{ii}$ is given by

$$[BA]_{ii} = \sum_{j=1}^{n} b_{ij} a_{ji}$$

Note that if we sum over the diagonal elements of $AB$ and $BA$, we get the same result. Hence, $\text{tr}(AB) = \text{tr}(BA)$.

**30.** Let $A$ be a diagonal matrix given by:

$$A = \begin{pmatrix} a_{11} & 0 & ... & 0 \\ 0 & a_{22} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & a_{nn} \end{pmatrix}$$

Then, it is very clearly observable that

$$A^k = \begin{pmatrix} a_{11}^k & 0 & ... & 0 \\ 0 & a_{22}^k & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & a_{nn}^k \end{pmatrix}$$

Proving this is just induction.

**34.** Let $A$ be a diagonal matrix such that $[A]_{ii} = a_{ii}$, such that all diagonal elements are non-zero. Then, it is not hard to see that the diagonal matrix given by $[A']_{ii} = a_{ii}^{-1}$ is the inverse of $A$. So, all such matrices $A$ are invertible.

## III, §2. Exercises

**2.** Let $V$ and $W$ be vector spaces, and let $T : V \to W$ be a linear map. Observe that

$$T(2O_V) = 2T(O_V)$$

But we know that $2O_V = O_V$, which means that

$$T(O_V) = 2T(O_V)$$

and by adding the inverse of $T(O_V)$ on both sides, we get

$$O_W = T(O_V)$$

which proves the claim.

**7.** Let $V$ and $W$ be vector spaces, and let $F : V \to W$ be a linear map. Define

$$U := \{v \in V : F(v) = O_W\}$$

First, observe that $F(O_V) = O_W$, so $O_V \in U$. Next, suppose $u$ and $v$ are in $U$. Then,

$$F(u+v) = F(u) + F(v) = O_W$$

which means that $u + v \in U$. Finally, if $c \in K$, then $F(cu) = cF(u) = cO_W = O_W$ which means that $cu \in U$. Thus, $U$ is a subspace of $V$.

**15.** Suppose that $w_1, w_2, ..., w_n$ are linearly independent elements of $W$, and suppose $F(v_i) = w_i$. Also, suppose that the equation

$$a_1v_1 + a_2v_2 + ... + a_nv_n = O_V$$

holds for some numbers $a_1, a_2, ..., a_n$ in $K$. Then, observe that

$$F(a_1v_1 + a_2v_2 + ... + a_nv_n) = F(O_V) = O_W$$

which means that

$$a_1F(v_1) + a_2F(v_2) + ... + a_nF(v_n) = O_W$$

However, the last equation implies that

$$a_1 = a_2 = a_3 = ... = a_n = 0$$

because $w_1, w_2, ..., w_n$ are linearly independent. Hence, $v_1, v_2, ..., v_n$ are linearly independent.

**16.** Let $v_0$ be an element of $V$ that does not lie in $W$. So, $F(v_0)$ is a non-zero real number. Now, let $v \in V$. If $v \in W$, then we can write $v = v + 0 \cdot v_0$, and we are done. So, suppose $v \notin W$. So, $F(v) \neq 0$. Now, we can write

$$F(v) = (F(v)F(v_0)^{-1})F(v_0) = cF(v_0)$$

where $c = F(v)F(v_0)^{-1}$. But, $F$ is a linear map, so we can write

$$F(v) = F(cv_0)$$

which means that

$$F(v - cv_0) = 0$$

and thus $v - cv_0 \in W$. So, let $w = v - cv_0$, and thus we can write

$$v = w + cv_0.$$

**17.** In exercise **14** we showed that $W$ is a subspace of $V$. Now, if $\{v_1, v_2, ..., v_n\}$ is a basis of $W$, then using exercise **16**, it follows that $\{v_0, v_1, v_2, ..., v_n\}$ is a basis of $V$.

## The Rank-Nullity Theorem

In this section, I will attempt a proof of the so called Rank-Nullity theorem, one of the central results of linear algebra. Its a relatively simple proof.

**Theorem(Rank-Nullity):** Suppose $V$ and $W$ are vector spaces, and let $L : V \to W$ be a linear map. Let Ker$L$ be the kernel of $L$, and let im$L$ be the image of $V$ under $L$. Then,

$$\dim V = \dim \mathrm{Ker}L + \dim \mathrm{im}L$$

*Proof:* Let $\dim \mathrm{Ker}L = q$ and $\dim \mathrm{im}L = s$. Note that if $\mathrm{im}L = \{O\}$, then there is nothing to prove. So, let us assume that $s > 0$. We first look at the case where the kernel of $L$ is not $\{O\}$.

Let $\{w_1, w_2, ..., w_s\}$ be a basis of im$L$, and let $\{u_1, u_2, ..., u_q\}$ be a basis of Ker$L$. Let $\{v_1, v_2, ..., v_s\}$ be elements of $V$ such that $L(v_i) = w_i$ for each $1 \leq i \leq s$. We will show that $\{v_1, v_2, ..., v_s, u_1, u_2, ..., u_q\}$ is a basis of $V$, proving the theorem for this case.

Let $v \in V$. Then, there are elements $a_1, a_2, ..., a_s$ in $K$ such that

$$L(v) = a_1w_1 + a_2w_2 + ... + a_sw_s$$

which could also be written as

$$L(v) = a_1L(v_1) + a_2L(v_2) + ... + a_sL(v_s)$$

and since $L$ is linear, it follows that

$$v - a_1v_1 - a_2v_2 - ... - a_sv_s \in \mathrm{Ker}L$$

which in addition means that there are numbers $b_1, b_2, ..., b_q$ in $K$ such that

$$v - a_1v_1 - a_2v_2 - ... - a_sv_s = b_1u_1 + b_2u_2 + ... + b_qu_q$$

and so it follows that every $v \in V$ can be written as a linear combination of $\{v_1, v_2, ..., v_s, u_1, u_2, ..., u_q\}$. Now, we will show that this set of vectors is also linearly independent.

Suppose that there are numbers $x_1, x_2, ..., x_s$ and $y_1, y_2, ..., y_q$ in $K$ such that

$$x_1v_1 + x_2v_2 + ... + x_sv_s + y_1u_1 + y_2u_2 + ... + y_qu_q = O$$

Applying $L$ to the above equation, we see that

$$x_1w_1 + x_2w_2 + ... + x_sw_s = O$$

which by our assumption implies that

$$x_1 = x_2 = ... = x_s = 0$$

and thus at the same time implies that

$$y_1 = y_2 = ... = y_q = 0.$$

This completes the prove for this case.

Now, suppose $\mathrm{Ker}L = \{O\}$. Then, suppose $\{w_1, w_2, ..., w_s\}$ is a basis of im$L$. Again, if $\{v_1, v_2, ..., v_s\}$ are in $V$ such that $L(v_i) = w_i$, it is not hard to show that $\{v_1, v_2, ..., v_s\}$ is a basis of $V$. So, the proof is even simpler in this case.

## III, §3. Exercises

**1.** Let $A$ and $B$ be a basis of $\mathbb{R}^2$, and let $F : \mathbb{R}^2 \to \mathbb{R}^n$ be linear. Then, by the Rank-Nullity theorem, we have

$$2 = \dim \mathrm{Ker}F + \dim \mathrm{im}F$$

Now, three cases are possible. If $\mathrm{Ker}F = \{O\}$, then observe that $F(A)$ and $F(B)$ form a basis of im$F$, and thus they are linearly independent. If Ker$F$ is not $\{O\}$, then it has a non-zero dimension. If $\dim \mathrm{Ker}F = 1$, it follows that $\dim \mathrm{im}F = 1$. If $\dim \mathrm{Ker}F = 2$, then $\mathrm{Ker}F = \mathbb{R}^2$, and thus $\mathrm{im}F = \{O\}$.

**3.** Define a linear map $F : \mathbb{R}^4 \to \mathbb{R}^2$ given by

$$F(x_1, x_2, x_3, x_4) = (x_1 + 2x_2, x_3 - 15x_4)$$

The problem now reduces to finding the dimension of Ker$F$. Observe that, im$F$ is $\mathbb{R}^2$, which is very clearly observable. So, by Rank-Nullity, it follows that $\dim \text{Ker} F = 2$.

**9.** **(a)** Suppose $f \in \text{Ker}L$. Then, it follows that

$$f'(x) = f(x)$$

for all $x \in \mathbb{R}$. Every solution of this equation is of the form $ce^x$, where $c \in \mathbb{R}$.
**(b)** If we rather have to find the kernel of the map $D - aI$, then it consists of all functions $f$ of the form $ce^{ax}$, where $c \in \mathbb{R}$.

**10.** **(a)** Define the map $L : K^n \to K$ by the formula

$$L(a_1, a_2, ..., a_n) = a_1 + a_2 + ... + a_n$$

It is not difficult to see that $L$ is linear. Now, the problem reduces to finding the dimension of Ker$F$. Observe that im$F = K$, and hence by Rank-Nullity, we see that $\dim \text{Ker} F = n - 1$.
**(b)** Consider the space of $n \times n$ matrices with entries in $K$. Let's denote this space by $M_n(K)$. Then, we have that

$$\dim M_n(K) = n^2$$

Now, we sum of the diagonal elements is just the **trace** ($\mathbf{tr} : M_n(K) \to K$) of the matrix, which is a linear map. So, the subspace of $M_n(K)$ such that the trace of a matrix in this subspace is 0 is just Ker **tr**. Applying Rank-Nullity, we see that

$$\dim \text{Ker } \mathbf{tr} = n^2 - 1$$

**11.** We have already shown earlier that **tr** is a linear map, and that

$$\mathbf{tr}(AB) = \mathbf{tr}(BA)$$

for any two square matrices $A$ and $B$.
**(c)** Suppose $B$ is invertible. Then,

$$\begin{aligned}\mathbf{tr}(B^{-1}AB) &= \mathbf{tr}(B^{-1}(AB)) \\ &= \mathbf{tr}((AB)B^{-1}) \\ &= \mathbf{tr}(A(BB^{-1})) \\ &= \mathbf{tr}(A)\end{aligned}$$

**(d)** Let $A$ and $B$ be in $M_n(K)$, and define

$$\langle A, B \rangle = \mathbf{tr}(AB)$$

First, it is clear that $\langle A, B \rangle = \langle B, A \rangle$. Second, it is also clear that $\langle A, B + C \rangle = \langle A, B \rangle + \langle A, C \rangle$. The third property regarding scalar multiplication is also clear. So, this is a scalar product.
**(e)** Consider the equation

$$AB - BA = I_n$$

where $A$ and $B$ are elements of $M_n(K)$. Taking the trace of the left hand side, we see that

$$\mathbf{tr}(AB - BA) = \mathbf{tr}(AB) - \mathbf{tr}(BA) = 0$$

However, the trace of $I_n$ is non-zero. So, this means that there are no matrices $A$, $B$ which satisfy this equation.

**12.** Consider $S$, the set of symmetric $n \times n$ matrices. Let's verify the vector space axioms for $S$ (assume that the entries are in a field $K$). In the following, $A$, $B$ and $C$ are any three symmetric matrices, $a$ and $b$ are any scalars in $K$.

(1) We have that $(A+B)+C = A+(B+C)$
(2) We have $A+B = B+A$
(3) The zero matrix $O$ satisfies $O+A = A+O = A$
(4) For every $A$, there is a $-A$ such that $A+(-A) = O = (-A)+A$
(5) We have $a(A+B) = aA+aB$
(6) We have $(a+b)A = aA+bA$
(7) We have $a(bA) = (ab)A$
(8) We have $1A = A$

and thus $S$ is indeed a vector space.

Let us try to determine the dimension of $S$. Since these matrices are symmetric about their diagonals, it follows that

$$\dim S = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

For example, when $n = 2$, a possible basis is:

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

and when $n = 3$, a possible basis is:

$$\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

**14.** Let $A$ be a matrix in $M_n(K)$. Observe that

$$(A+A^t)^t = A^t + A = A + A^t$$

which means that $A + A^t$ is symmetric. Also,

$$(A-A^t)^t = A^t - A = -(A-A^t)$$

which means that $A - A^t$ is skew symmetric. Finally, write

$$A = \frac{A+A^t}{2} + \frac{A-A^t}{2}$$

We can also show that this representation is unique. Let

$$A = B + C = B_1 + C_1$$

where both $B$ and $B_1$ are symmetric, and $C$ and $C_1$ are skew symmetric. So, we have

$$B - B_1 = C_1 - C$$

Take the transpose of both sides to get

$$(B-B_1)^t = (C_1 - C)^t$$

which means that

$$B - B_1 = C_1^t - C^t = C - C_1$$

which means that

$$2(B - B_1) = O$$

which implies that $B = B_1$. And from there, we can conclude that $C = C_1$.

**15.** Define $P : M_n(K) \to M_n(K)$ by the formula

$$P(A) = \frac{A + A^t}{2}$$

**(a)** Let $A$, $B$ be in $M_n(K)$ and let $c \in K$. Then,

$$P(A+B) = \frac{A + B + (A+B)^t}{2} = \frac{A + A^t}{2} + \frac{B + B^t}{2} = P(A) + P(B)$$

and

$$P(cA) = \frac{cA + (cA)^t}{2} = \frac{cA + cA^t}{2} = c\frac{A + A^t}{2} = cP(A)$$

which means that $P$ is linear.

**(b)** Suppose $P(S) = O$ for some $S \in M_n(K)$. This means that

$$\frac{S + S^t}{2} = O$$

which implies that

$$S = -S^t$$

and hence $\mathrm{Ker}P$ is the space of skew-symmetric matrices in $M_n(K)$.

**(c)** Now, $\frac{A+A^t}{2}$ is always a symmetric matrix, and hence $\mathrm{im}P$ is the subspace of all symmetric matrices in $M_n(K)$, whose dimension is $\frac{n(n+1)}{2}$. Applying Rank-Nullity, we see that

$$\dim \mathrm{Ker}P = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

or in simple words, the dimension of the subspace of skew-symmetric matrices in $M_n(K)$ is $\frac{n(n-1)}{2}$.

**17.** **(b)** Let $U$ and $W$ be two vector spaces of dimension $n$ and $m$ respectively. It is easy to see that the dimension of $U \times W$ is $n + m$. To see this, let $\{u_1, u_2, ..., u_n\}$ be a basis of $U$, and let $\{w_1, w_2, ..., w_m\}$ be a basis of $W$. Now, any element $(u, w) \in U \times W$ can be written as

$$(u, w) = a_1(u_1, 0) + a_2(u_2, 0) + ... + a_n(u_n, 0) + b_1(0, w_1) + b_2(0, w_2) + ... + b_m(0, w_m)$$

## III, §4. Exercises

**1.** Suppose $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map such that $L \neq O$ and $L^2 = O$. Since $L \neq O$, there is some $A \in \mathbb{R}^2$ such that

$$L(A) = B$$

for some $B \neq O$. It is clear that

$$L(B) = L(L(A)) = O.$$

We will show that $A$ and $B$ are linearly independent, thus implying that they form a basis of $R^2$.

Suppose $x$ and $y$ are real numbers such that

$$xA + yB = O.$$

Applying $L$ to both sides, we get

$$xL(A) + yL(B) = O$$

But since $L(B) = O$, it means that

$$xL(A) = O$$

which means that $x = 0$(because $L(A) = B \neq O$). Also, it means that $y = 0$. Hence, $A$ and $B$ are linearly independent, meaning that they form a basis of $R^2$.

**10.** Let $v \in V$. Write $v$ as

$$v = v - P(v) + P(v)$$

Observe that

$$P(v - P(v)) = P(v) - P(P(v)) = P(v) - P(v) = O$$

which means that $v - P(v) \in \mathrm{Ker}P$. Next, let $x \in \mathrm{Ker}P \cap \mathrm{im}P$. Then,

$$P(x) = O \quad \text{and} \quad x = P(y)$$

for some $y \in V$. But, this means that

$$P(P(y)) = P(y) = P(x) = O = x$$

and thus $\mathrm{Ker}P \cap \mathrm{im}P = \{O\}$.

**16.** Let $V$ and $W$ be vector spaces over a field $K$, both of dimension $n$. For any $v \in V$, we can find numbers $a_1, a_2, ..., a_n$ in $K$ such that

$$v = a_1v_1 + a_2v_2 + ... + a_nv_n$$

and these numbers are unique. So, consider the map $L : V \to W$ given by:

$$L(v) = a_1w_1 + a_2w_2 + ... + a_nw_n$$

It is clear that $L$ is linear. We will show that $L$ is an isomorphism. First, let $x \in \mathrm{Ker}L$. Then,

$$L(x) = O$$

If $x = b_1v_1 + b_2v_2 + ... + b_nv_n$, this means that

$$L(x) = b_1w_1 + b_2w_2 + ... + b_nw_n = O$$

which means that $b_1 = b_2 = ... = b_n = 0$, since $w_1, w_2, ..., w_n$ are linearly independent. So, $\mathrm{Ker}L = \{O\}$.

By the Rank-Nullity theorem, it follows that

$$\dim \mathrm{Im}L = \dim V = \dim W = n$$

which means that $\mathrm{Im}L = W$.

Hence, $L$ is both injective and surjective, and hence is bijective. So, its an isomorphism, and hence $V$ and $W$ are isomorphic.

## Matrices and Linear Maps

In this section, I will prove that every linear map can be thought of as a matrix. Let's start with a special case.

**Theorem:** Let $K$ be a field, and let $L : K^n \to K^m$ be a linear map. Then, there is a unique $m \times n$ matrix $A$ such that

$$L(X) = AX$$

for any column vector $X$ in $K^n$. Here, $AX$ is the matrix product of $A$ and $X$.

*Proof:* Let $E_1, E_2, ..., E_n$ be a basis of $K^n$, and let $e_1, e_2, ..., e_m$ be a basis of $K^m$. Note that the values of $L(E_1), L(E_2), ..., L(E_n)$ determine $L$. So, suppose

$$\begin{aligned} L(E_1) &= a_{11}e_1 + a_{21}e_2 + ... + a_{m1}e_m \\ L(E_2) &= a_{12}e_1 + a_{22}e_2 + ... + a_{m2}e_m \\ &........ \\ L(E_n) &= a_{1n}e_1 + a_{2n}e_2 + ... + a_{mn}e_m \end{aligned}$$

So, if $X = x_1E_1 + x_2E_2 + ... + x_nE_n$ then

$$\begin{aligned} L(X) =& x_1(a_{11}e_1 + a_{21}e_2 + ... + a_{m1}e_m)+ \\ & x_2(a_{12}e_1 + a_{22}e_2 + ... + a_{m2}e_m)+ \\ & ...... \\ & +x_n(a_{1n}e_1 + a_{2n}e_2 + ... + a_{mn}e_m) \end{aligned}$$

and writing in component form, we get:

$$\begin{aligned} L(X) =& e_1(a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n)+ \\ & e_2(a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n)+ \\ & e_m(a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n)+ \end{aligned}$$

or, in matrix multiplication form, we can write

$$\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{12} & ... & a_{1n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 & a_{12}x_2 & ... & a_{1n}x_n \\ a_{21}x_1 & a_{12}x_2 & ... & a_{1n}x_n \\ ... & ... & ... & ... \\ a_{m1}x_1 & a_{m2}x_2 & ... & a_{mn}x_n \end{pmatrix} = L(X)$$

The uniqeness of $A$ is very easy to prove.

**NOTE:** This proof also gives us a quick way to find the matrix of such a linear map. Find numbers $a_{ij}$ in $K$ such that

$$\begin{aligned} L(E_1) &= a_{11}e_1 + a_{21}e_2 + ... + a_{m1}e_m \\ L(E_2) &= a_{12}e_1 + a_{22}e_2 + ... + a_{m2}e_m \\ &........ \\ L(E_n) &= a_{1n}e_1 + a_{2n}e_2 + ... + a_{mn}e_m \end{aligned}$$

and the transpose of the matrix

$$\begin{pmatrix} a_{11} & a_{21} & ... & a_{m1} \\ a_{12} & a_{22} & ... & a_{m2} \\ ... & ... & ... & ... \\ a_{1n} & a_{2n} & ... & a_{mn} \end{pmatrix}$$

will be the matrix associated with the map $L$.

## IV, §2. Exercises

These exercises will be a very good practice to get familiar with how linear maps and matrices are related to each other.

**1.** In these computations, the general method remains that same: find the co-ordinates of the values at basis elements, and take the inverse of the coordinate matrix to get the matrix of the linear map!
**(a)** $F : \mathbb{R}^4 \to \mathbb{R}^2$ given by $F({}^t(x_1, x_2, x_3, x_4)) =^t (x_1, x_2)$: Here, let's compute $F$ at the basis elements:

$$F({}^t(1,0,0,0)) =^t (1,0)$$
$$F({}^t(0,1,0,0)) =^t (0,1)$$
$$F({}^t(0,0,1,0)) =^t (0,0)$$
$$F({}^t(0,0,0,1)) =^t (0,0)$$

and thus the matrix associated with $F$ is:

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

**(c)** $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $F({}^t(x,y)) =^t (3x, 3y)$. Again, $F$ at the basis elements is given by:

$$F({}^t(1,0)) =^t (3,0)$$
$$F({}^t(0,1)) =^t (0,3)$$

and so matrix is:

$$\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

**(d)** $F : \mathbb{R}^n \to \mathbb{R}^n$ given by $F(X) = 7X$. Here, $F$ at the basis elements will look like:

$$F(1,0,0,...,0) = (7,0,0,...,0)$$
$$F(0,1,0,...,0) = (0,7,0,...,0)$$
$$....$$
$$F(0,0,0,...,1) = (0,0,0,...,7)$$

and thus the matrix of $F$ will be given by:

$$\begin{pmatrix} 7 & 0 & 0 & ... & 0 \\ 0 & 7 & 0 & ... & 0 \\ 0 & 0 & 7 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 7 \end{pmatrix}$$

which is the diagonal matrix with the all diagonal elements equal to 7.

**(f)** $F : \mathbb{R}^4 \to \mathbb{R}^2$ given by $F(^t(x_1, x_2, x_3, x_4)) =^t (x_1, x_2, 0, 0)$. $F$ at the basis elements is given by:

$$\begin{aligned} F(1,0,0,0) &= (1,0,0,0) \\ F(0,1,0,0) &= (0,1,0,0) \\ F(0,0,1,0) &= (0,0,0,0) \\ F(0,0,0,1) &= (0,0,0,0) \end{aligned}$$

and thus the matrix is:

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

## Matrices and General Linear Maps

In this section I'll not prove anything, but I will discuss how we can relate any general linear map to a matrix. Let $V$, $W$ be vector spaces over some field $K$, and let $L : V \to W$ be a linear map. In addition, suppose $\mathscr{R} = \{v_1, v_2, ..., v_n\}$ is a basis of $V$, and suppose $\mathscr{R}' = \{w_1, w_2, ..., w_m\}$ be a basis of $W$. Note that the map $L$ is completely determined by the values of $L(v_1), L(v_2), ..., L(v_n)$. So, suppose there are numbers $a_{ij}$ in $K$ such that

$$\begin{aligned} L(v_1) &= a_{11}w_1 + a_{21}w_2 + ... + a_{m1}w_m \\ L(v_2) &= a_{12}w_1 + a_{22}w_2 + ... + a_{m2}w_m \\ &.... \\ L(v_n) &= a_{1n}w_1 + a_{2n}w_2 + ... + a_{mn}w_m \end{aligned}$$

Now, suppose $v \in V$. Then, it has *coordinates* with respect to the basis $\mathscr{R}$. Let the *coordinate vector* of $v$ be denoted by $X_{\mathscr{R}}(v)$. For example, if $v = v_1 + v_2 + ... + v_n$, then $X_{\mathscr{R}}(v) = (1, 1, 1, ..., 1)$. Note that $X_{\mathscr{R}}(v) \in K^n$. So, using coordinate vectors, we can say that

$$X_{\mathscr{R}'}(L(v)) = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix} X_{\mathscr{R}}(v) = M^{\mathscr{R}}_{\mathscr{R}'}(L) X_{\mathscr{R}}(v)$$

Note that the matrix $M^{\mathscr{R}}_{\mathscr{R}'}(L)$ is the transpose of the matrix

$$\begin{pmatrix} a_{11} & a_{21} & ... & a_{m1} \\ a_{12} & a_{22} & ... & a_{m2} \\ ... & ... & ... & ... \\ a_{1n} & a_{2n} & ... & a_{mn} \end{pmatrix}.$$

This is very similar to what we did for maps from $K^n$ to $K^m$.

We can also say that the space $\mathscr{L}(V, W)$ of all linear maps from $V$ to $W$ is isomorphic to the space of all $m \times n$ matrices over the field $K$ under the isomorphism

$$f \mapsto M^{\mathscr{R}}_{\mathscr{R}'}(f)$$

This gives us a very powerful formula to describe any linear map. For instance, let's talk about converting from one basis to another. If $\mathscr{R}$ and $\mathscr{R}'$ are two basis of a vector space $V$, then the map associated with the matrix $M^{\mathscr{R}}_{\mathscr{R}'}(\text{id})$ converts a vector represented in basis $\mathscr{R}$ to the same vector represented in basis $\mathscr{R}'$. Also, observe that

$$M^{\mathscr{R}}_{\mathscr{R}'}(\text{id})M^{\mathscr{R}'}_{\mathscr{R}}(\text{id}) = I_n = M^{\mathscr{R}'}_{\mathscr{R}}(\text{id})M^{\mathscr{R}}_{\mathscr{R}'}(\text{id})$$

where $I_n$ is the $n \times n$ identity matrix over the field $K$, and id is the identity map.

## IV, §3. Exercises

**1.** **(a)** Here, we have $\mathscr{R} = \{(1,1,0),(-1,1,1),(0,1,2)\}$ and $\mathscr{R}' = \{(2,1,1),(0,0,1),(-1,1,1)\}$ Now, we have

$$\begin{aligned}
(1,1,0) &= \frac{2}{3}(2,1,1) - (0,0,1) + \frac{1}{3}(-1,1,1)\\
(-1,1,1) &= 0(2,1,1) + 0(0,0,1) + (-1,1,1)\\
(0,1,2) &= \frac{1}{3}(2,1,1) + (0,0,1) + \frac{2}{3}(-1,1,1)
\end{aligned}$$

and thus

$$M^{\mathscr{R}}_{\mathscr{R}'}(\text{id}) = \frac{1}{3}\begin{pmatrix} 2 & 0 & 1\\ -3 & 0 & 3\\ 1 & 3 & 2\end{pmatrix}$$

**(b)** Here we have $\mathscr{R} = \{(3,2,1),(0,-2,5),(1,1,2)\}$ and $\mathscr{R}' = \{(1,1,0),(-1,2,4),(2,-1,1)\}$. Now, we have

$$\begin{aligned}
(3,2,1) &= \frac{11}{5}(1,1,0) + \frac{2}{15}(-1,2,4) + \frac{7}{15}(2,-1,1)\\
(0,-2,5) &= \frac{-11}{5}(1,1,0) + \frac{13}{15}(-1,2,4) + \frac{23}{15}(2,-1,1)\\
(1,1,2) &= \frac{3}{5}(1,1,0) + \frac{2}{5}(-1,2,4) + \frac{2}{5}(2,-1,1)
\end{aligned}$$

and thus

$$M^{\mathscr{R}}_{\mathscr{R}'}(\text{id}) = \frac{1}{15}\begin{pmatrix} 33 & -33 & 9\\ 2 & 13 & 6\\ 7 & 23 & 6\end{pmatrix}$$

**2.** Suppose $L : V \to V$ is a linear map. Let $\mathscr{R} = \{v_1, v_2, ..., v_n\}$ be a basis of $V$, such that $L(v_i) = c_i v_i$. Then,

$$M^{\mathscr{R}}_{\mathscr{R}} = \begin{pmatrix} c_1 & 0 & 0 & 0 & .... & 0\\ 0 & c_2 & 0 & 0 & .... & 0\\ 0 & 0 & c_3 & 0 & .... & 0\\ 0 & 0 & 0 & c_4 & .... & 0\\ ... & ... & ... & ... & .... & ...\\ 0 & 0 & 0 & 0 & .... & c_n\end{pmatrix}$$

which means that $L$ is diagonalisable.

**7.** Let $F$ be the rotation through an angle $\theta$. Let $\{E_1, E_2\}$ be the usual basis of $\mathbb{R}^2$. Now, the rotated coordinate system has basis $F(E_1)$ and $F(E_2)$, which are given by:

$$F(E_1) = (\cos\theta, \sin\theta)$$
$$F(E_2) = (-\sin\theta, \cos\theta)$$

Now, if $(x, y)$ are the coordinates of a point w.r.t the usual basis $\mathscr{R} = \{E_1, E_2\}$, we wish to determine the coordinates w.r.t the basis $\mathscr{R}' = \{F(E_1), F(E_2)\}$, which is the rotated coordinate system.

Now, we have

$$(1, 0) = \cos\theta(\cos\theta, \sin\theta) - \sin\theta(-\sin\theta, \cos\theta)$$
$$(0, 1) = \sin\theta(\cos\theta, \sin\theta) + \cos\theta(-\sin\theta, \cos\theta)$$

and so the matrix $M_{\mathscr{R}'}^{\mathscr{R}}(\mathrm{id})$ is given by

$$M_{\mathscr{R}'}^{\mathscr{R}}(\mathrm{id}) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and so, the coordinates of the point $(x, y)$ w.r.t the rotated system is given by:

$${}^t(x', y') = M_{\mathscr{R}'}^{\mathscr{R}}(\mathrm{id})\,{}^t(x, y) = {}^t\,(x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta)$$

**8.** In this exercise I will only describe how to do part **(a)**, and the other parts have the same method.

**(a)** Suppose $\mathscr{R} = \{e^t, e^{2t}\}$. Then, $V$ is the vector space of all linear combinations of these two functions. Let $D : V \to V$ be the derivative. We only need to find $D$ at the basis elements. So, observe that

$$D(e^t) = e^t + 0e^{2t}$$
$$D(e^{2t}) = 0e^t + 2e^{2t}$$

and hence

$$M_{\mathscr{R}}^{\mathscr{R}}(D) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

**(b)** We have

$$D(1) = 0(1) + 0t + 0e^t + 0e^{2t} + 0te^{2t}$$
$$D(t) = 1 + 0t + 0e^t + 0e^{2t} + 0te^{2t}$$
$$D(e^t) = 0(1) + 0t + e^t + 0e^{2t} + 0te^{2t}$$
$$D(e^{2t}) = 0(1) + 0t + 0e^t + 2e^{2t} + 0te^{2t}$$
$$D(te^{2t}) = 0(1) + 0t + 0e^t + e^{2t} + 2te^{2t}$$

and hence

$$M_{\mathscr{R}}^{\mathscr{R}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

**10.** Let $P_n$ be the vector space of all polynomials of degree $\leq n$, and let $D : \mathbb{P}_n \to \mathbb{P}_n$ be the derivative. Observe that, the map $D^n$ is the $O$ map, because the $n^{\text{th}}$ derivative of a polynomial in $P_n$ must be the zero polynomial. This means that $D$ is actually nilpotent.

**(a)** Observe that

$$\begin{aligned}
&(I - D^2) \circ (I + D^2 + D^4 + ... + D^{2n-2}) \\
&\qquad = I \circ (I + D^2 + D^4 + ... + D^{2n-2}) - D^2 \circ (I + D^2 + D^4 + ... + D^{2n-2}) \\
&\qquad = I + D^2 + ... + D^{2n-2} - D^2 - D^4 - ... - D^{2n} \\
&\qquad = I - D^{2n} \\
&\qquad = I
\end{aligned}$$

because $D^{2n}$ is also the $O$ map.

**(b)** We can extend this technique to a general map $D^m - I$. Here, observe that

$$(D^m - I) \circ (-I - D^m - D^{2m} - ... - D^{(n-1)m}) = I - D^{mn} = I$$

because $D^{mn}$ is the $O$ map.

## Scalar Products

Let $V$ be a vector space over a field $K$. Then, the scalar product is simply a function from that takes two vectors and returns a scalar. I won't list the axioms here. For a vector space over $\mathbb{R}$, we say that a scalar product is *positive definite* if for every $v \in V$,

$$\langle v, v \rangle \geq 0$$

and if $v \neq O$, then the inequality is strict. Let's prove two distance theorems:

**Pythagoras Theorem:** If $v$ and $w$ are perpendicular, then

$$||v + w||^2 = ||v||^2 + ||w||^2$$

*Proof:* We have

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

**Parallelogram Law:** For any $v, w$ in $V$, we have

$$||v + w||^2 + ||v - w||^2 = 2||v||^2 + 2||w||^2$$

*Proof:* We have

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

There are two useful inequalities associated with norms and the scalar product (I am not proving them here):

**Cauchy-Schwarz Inequality:** If $V$ is a vector space with a positive-definite scalar product, then

$$|\langle v, w \rangle| \leq ||v||||w||$$

**Triangle Inequality:** If $V$ is a vector space with a positive-definite scalar product, then

$$||v+w|| \leq ||v|| + ||w||$$

## V,§1. Exercises

**1.** Suppose $V$ is a vector space with a scalar product. Then, for any $v \in V$

$$\langle O, v\rangle = \langle O+O, v\rangle = \langle O, v\rangle + \langle O, v\rangle$$

which implies that $\langle O, v\rangle = 0$.

**2.** Suppose that $V$ has a positive definite scalar product. Let $v_1, v_2, ..., v_n$ be non-zero elements that are mutually perpendicular. We will show that these vectors are linearly independent. So, let $a_1, a_2, ..., a_n$ be in $K$ such that

$$a_1v_1 + a_2v_2 + ... + a_nv_n = O$$

Then, taking the scalar product of both sides with $v_1$, we get

$$\langle a_1v_1 + ... + a_nv_n, v_1\rangle = \langle O, v_1\rangle = 0$$

Also, we have

$$\langle a_1v_1 + ... + a_nv_n, v\rangle = a_1\langle v_1, v_1\rangle$$

which means that

$$a_1\langle v_1, v_1\rangle = O$$

Since the scalar product is positive definite and $v_1 \neq O$, this means that $a_1 = 0$. Similarly, it can be proved that

$$a_1 = a_2 = ... = a_n = 0$$

and thus the vectors are linearly independent.

**3.** Let $M$ be a symmetric $n \times n$ matrix. Let $X$ and $Y$ be column $n$-vectors, and define

$$\langle X, Y\rangle = X^tMY$$

Since $X^tMY$ is a $1 \times 1$ matrix, it is symmetric, and thus

$$(X^tMY)^t = X^tMY$$

which means that

$$Y^tM^tX = Y^tMX = X^tMY$$

and so $\langle X, Y\rangle = \langle Y, X\rangle$. The other two properties may also be easily verified, and hence this is a scalar product.

## Orthogonal and Orthonormal Basis

In this section, we will prove a very important theorem about the existence of orthogonal basis.

**Theorem 0.8.** Suppose $V$ is an $n$-dimensional vector space over $\mathbb{R}$ with a positive definite scalar product. Suppose $W$ is a subspace of $V$, and suppose $\{w_1, w_2, ..., w_m\}$ is an orthogonal basis of $W$. If $W \neq V$, then there exist elements $w_{m+1}, ..., w_n$ such that $\{w_1, w_2, ..., w_m, w_{m+1}, .., w_n\}$ is an orthogonal basis of $V$.

Note: In a way, this theorem is saying that if we start with a set of linearly independent vectors which are also orthogonal to each other, we can always extend them to form a basis of $V$.

*Proof:* We know that we can find elements $v_{m+1}, ..., v_n$ such that

$$\{w_1, w_2, ..., w_m, v_{m+1}, ..., v_n\}$$

is a basis of $V$. But, this basis may not be orthogonal. The idea is to replace $v_{m+1}$ with some $w_{m+1}$. Consider the vector

$$w_{m+1} = v_{m+1} - c_1 w_1 - c_2 w_2 - ... - c_m w_m$$

where $c_i = \dfrac{\langle v_{m+1}, w_i\rangle}{\langle w_i, w_i\rangle}$, i.e $c_i$ is the component of $v_{m+1}$ along $w_i$. Note that, positive definiteness is required for each $c_i$ to we well defined. Now, it is clear that $w_{m+1}$ is perpendicular to every $w_i$. Also, we have

$$v_{m+1} = w_{m+1} + c_1 w_1 + ... + c_m w_m$$

which means that $\{w_1, w_2, ..., w_m, w_{m+1}, v_{m+2}, .., v_n\}$ is a basis of $V$. We repeat this step for $v_{m+2}$, $v_{m+3}$ till $v_n$. Thus, we obtain an orthogonal basis.

A corollary of this theorem is that every finite dimensional vector space $V$ that is not equal to $\{O\}$ and which has a positive definite scalar product has an orthogonal basis. This proof also tells us how to construct one.

Given an orthogonal basis, we can find an orthonormal basis by dividing each vector by its norm.

**Exercise:** As an exercise let's find an orthonormal basis for the vector space generated by the vectors $v_1 = (1, 1, 0, 1)$, $v_2 = (1, -2, 0, 0)$ and $v_3 = (1, 0, -1, 2)$. These vectors are linearly independent. Now, orthogonalise the second vector w.r.t the first and get

$$v_2' = v_2 - \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1 = \left(\frac{4}{3}, \frac{-5}{3}, 0, \frac{1}{3}\right)$$

Then, orthogonalise $v_3$ w.r.t $v_1$ and $v_2'$ to obtain

$$v_3' = v_3 - \frac{\langle v_3, v_1\rangle}{\langle v_1, v_1\rangle} v_1 - \frac{\langle v_3, v_2'\rangle}{\langle v_2', v_2'\rangle} v_2' = \left(\frac{-4}{7}, \frac{-2}{7}, -1, \frac{6}{7}\right)$$

So, the vectors $v_1, v_2', v_3'$ are an orthogonal basis. To get an orthonormal basis, divide these vectors by their norms.

The next theorem gives is both theoretically and practically powerful:

**Theorem 0.9.** Suppose $V$ is a vector space over $\mathbb{R}$ with a positive definite scalar product, and let $V$ have dimension $n$. Let $W$ be a subspace of $V$ of dimension $r$. Denote by $W^\perp$ the space of vectors perpendicular to $W$. Then, $V$ is a direct sum of $W$ and $W^\perp$, and $W^\perp$ has dimension $n - r$.

*Proof:* If $W = \{O\}$, then $W^\perp = V$, and the theorem is true. If $W = V$, then $W^\perp = \{O\}$, and again the theorem is true. So, let's assume that $W$ is not one of those subspaces.

There exists an orthogonal basis of $W$; let it be $\{w_1, w_2, ..., w_r\}$. Then, by a previous theorem, there are elements $u_{r+1}, ..., u_n$ of $V$ such that

$$\{w_1, w_2, ..., w_r, u_{r+1}, ..., u_n\}$$

is an orthogonal basis of $V$. We will show that $\{u_{r+1}, u_{r+2}, ..., u_n\}$ is an orgthogonal basis of $W^\perp$.

Suppose $w \in W^\perp$. Then, we can write $w$ as

$$w = a_1 w_1 + a_2 w_2 + ... + a_r w_r + a_{r+1} u_{r+1} + ... + a_n u_n$$

Taking the scalar product of $w$ with any $w_i$, we obtain

$$\langle w, w_i \rangle = 0 = a_i \langle w_i, w_i \rangle$$

which implies that $a_i = 0$. So, $u_{r+1}, ..., u_n$ generate $W^\perp$. Also, these elements are mutually perpendicular and linearly independent, so they constitute an orthogonal basis of $W^\perp$. Also, it is clear that $V$ is a direct sum of $W$ and $W^\perp$.

The importance of this theorem is that given any subspace, we can find the dimension of the space that is orthogonal to this space. $W^\perp$ is also called the **orthogonal complement** of $W$.

## V, §2. Exercises

**0.** Let $V = \mathbb{R}^6$ and $W$ be the space generated by the vectors $(1, 1, -2, 3, 4, 5)$ and $(0, 0, 1, 1, 0, 7)$. Clearly, $\dim W = 2$. We are interested in finding the dimension of $W^\perp$. We know that

$$\dim V = \dim W + \dim W^\perp$$

and it follows that $\dim W^\perp = 4$.

**1.** We will only do one part of this problem.
**(b)** Consider the subspace of $\mathbb{R}^3$ spanned by the vectors $A = (2, 1, 1)$ and $B = (1, 3, -1)$. One basis of this subspace is obviously $\{A, B\}$. However, $A$ and $B$ are not perpendicular. So, let's orthogonalise $B$ with respect to $A$ to obtain

$$B' = B - \frac{B \cdot A}{A \cdot A} A = \left( \frac{-1}{3}, \frac{7}{3}, \frac{-5}{3} \right)$$

and thus $\{A, B'\}$ is an orthogonal basis of the subspace. To get an orthonormal basis, we just divide these vectors by their norm, and so an orthonormal basis is

$$\left\{ \frac{1}{\sqrt{6}}(2, 1, 1), \frac{1}{\sqrt{75}}(-1, 7, -5) \right\}$$

**3.** For real-valued continuous functions on $[0, 1]$, let's define

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt$$

It is clear that

$$\langle f, g \rangle = \langle g, f \rangle$$

Also, we have

$$\int_0^1 (f(t) + g(t))h(t)dt = \int_0^1 f(t)h(t)dt + \int_0^1 g(t)h(t)dt$$

which gives us

$$\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$$

Finally,

$$\langle cf, g \rangle = \int_0^1 cf(t)g(t)dt = c \int_0^1 f(t)g(t)dt = c\langle f, g \rangle$$

and thus this is a well-defined scalar product. Also, it is not hard to see that this scalar product is actually positive definite and non-degenerate.

**4.** Let $V$ be the subspace of functions generated by $f(t) = t$ and $g(t) = t^2$. It is then clear that one basis of $V$ is

$$\{f, g\}$$

However, we have

$$\langle f, g\rangle = \int_0^1 t^3 dt \neq 0$$

and hence this is not an orthonormal basis. So, let's orthogonalise $g$ with respect to $f$ and get

$$g'(t) = g(t) - \frac{\langle g, f\rangle}{\langle f, f\rangle} f(t) = t^2 - \frac{3t}{4}$$

($g'$ is not the derivative!) and hence $\{g', f\}$ is an orthogonal basis of $V$. On dividing by their norms, we observe that an orthonormal basis is

$$\left\{\sqrt{3}t, \sqrt{80}\left(t^2 - \frac{3t}{4}\right)\right\}$$

**5.** Let $V$ be the space of functions generated by $f(t) = 1$, $g(t) = t$ and $h(t) = t^2$. An obvious basis is given by

$$\{f, g, h\}$$

Now we will do the exact same method as we did in **4.** So,

$$g'(t) = g(t) - \frac{\langle g, f\rangle}{\langle f, f\rangle} f(t) = t - \frac{1}{2}$$

and

$$h'(t) = h(t) - \frac{\langle h, f\rangle}{\langle f, f\rangle} f(t) - \frac{\langle h, g'\rangle}{\langle g', g'\rangle} g'(t) = t^2 - \frac{1}{3} - \left(t - \frac{1}{2}\right) = t^2 - t + \frac{1}{6}$$

and so

$$\{f, g', h'\}$$

is an orthogonal basis of $V$. To get an orthonormal basis, we just divide by the norm of these functions.

**6.** In this exercise, we will use the hermitian inner product.
(a) Let $V$ be the space generated by $A = (1, i, 0)$ and $B = (1, 1, 1)$. So, a clear cut basis of $V$ is

$$\{A, B\}$$

However, this is not an orthonormal basis. Let's orthogonalise $B$ with respect to $A$ to get

$$B' = B - \frac{\langle B, A\rangle}{\langle A, A\rangle} A = \left(\frac{1+i}{2}, \frac{1-i}{2}, 1\right)$$

Now, observe that $\langle A, B'\rangle = \langle B', A\rangle = 0$, and hence $\{A, B'\}$ is an orthogonal basis of $V$. So, an orthonormal basis is of $V$ is

$$\left\{\left(\frac{1}{\sqrt{2}}, \frac{i}{\sqrt{2}}, 0\right), \left(\frac{1+i}{2\sqrt{2}}, \frac{1-i}{2\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\right\}$$

**(b)** Let $V$ be the complex space generated by $A = (1, -1, -i)$ and $B = (i, 1, 2)$. So, a basis of $V$ is clearly

$$\{A, B\}$$

but this basis is clearly not orthogonal. Orthogonalising $B$ with respect to $A$, we get

$$B' = B - \frac{\langle B, A\rangle}{\langle A, A\rangle}A = \left(\frac{1}{3}, \frac{2+3i}{3}, \frac{3-i}{3}\right)$$

and so $\{A, B'\}$ is an orthogonal basis of $V$. So, an orthonormal basis is given by:

$$\left\{\left(\frac{1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}, \frac{-i}{\sqrt{3}}\right), \left(\frac{1}{2\sqrt{6}}, \frac{2+3i}{2\sqrt{6}}, \frac{3-i}{2\sqrt{6}}\right)\right\}$$

**7.**

**(a)** We have already shown that for any $n \times n$ matrices $A$ and $B$, it is true that

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$
$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

and hence this is a scalar product. To show that it is also non-degenerate, suppose there is a matrix $M$ such that

$$\mathrm{tr}(MX) = 0$$

for all square matrices $X$. If $M_i$ is the $i^{\text{th}}$ row of $M$, this means that

$$M_1 \cdot X^1 + M_2 \cdot X^2 + ... + M_n \cdot X^n = 0$$

where each $X^i$ is a column vector with $n$ elements. So, this means that every entry of $M$ is 0, and hence $M = O$. So, the trace is a non-degenerate scalar product.

**(b)** Suppose $A$ is a symmetric square matrix. Let $A_i$ be the $i^{\text{th}}$ row vector of $A$, and let $A^i$ be the $i^{\text{th}}$ column vector of $A$. Now, we have

$$\mathrm{tr}(A^2) = \sum_{i=1}^{n} A_i \cdot A^i$$

Since $A$ is symmetric, we have that $A_i = A^i$. So, the trace is $\geq 0$, and if $A \neq O$, then the trace is strictly positive.

**(c)** The dimension of the space of symmetric $n \times n$ matrices is $\dfrac{n(n+1)}{2}$ as we proved earlier. Now, observe that if $A$ is a matrix whose trace is 0, then it means that

$$a_{11} + a_{22} + ... + a_{nn} = 0$$

Given any values of $a_{11}$, $a_{22}$ upto $a_{(n-1)(n-1)}$, a value of $a_{nn}$ can be determined which makes the trace 0. So, the dimension of $W$ is

$$\dim W = \dim V - 1$$

and hence the dimension of the orthogonal complement of $W$ is 1.

## Proof that Column Rank = Row Rank

Let $A$ be an $m \times n$ matrix. The 'column' rank of $A$ is the dimension of the space generated by the columns of $A$. Similar is the definition of the 'row' rank. Now, we will prove that the column rank is equal to the row rank for a matrix over a field $K$, where in this discussion $K$ is either $\mathbb{R}$ or $\mathbb{C}$. Before giving a proof, we will make use of the fact that if $W$ is a subspace of $K^n$, then

$$\dim W + \dim W^{\perp} = n$$

Here, the associated scalar product is the normal dot $\cdot$ product. Note in this case of $\mathbb{C}$, the normal dot product is not positive definite. But, as we will prove later, this theorem still remains true. For $\mathbb{R}$, we have already proven this theorem because the dot product is positive definite.

If $A$ is a matrix over $K$, let $L_A$ be the linear map associated with it. So, $L_A : K^n \to K^m$ is a linear map. We are interested with the kernel of $L_A$. By Rank-Nullity, we have

$$\dim \mathrm{Ker} L_A + \dim \mathrm{Im} L_A = n$$

And the key observation is that

$$\text{Column Rank of A} = \dim \mathrm{Im} L_A$$

Now, we can interpret $\mathrm{Ker}L_A$ in another way: if $X \in \mathrm{Ker}L_A$, then

$$AX = O$$

which means that the dot product of $X$ with every row of $A$ is 0. In other words, if $Q$ is the space generated by the rows of $A$, then $\mathrm{Ker}L_A$ is the orthogonal complement of $Q$ with respect to the dot product. So, it follows that

$$\dim Q + \dim KerL_A = n$$

But, Row Rank of A $= \dim Q$, and hence it follows that

$$\text{Row Rank of A} = \text{Column Rank o A}$$

Additionally, the dimension of the kernel of $L_A$ is $n - \mathrm{Rank}(A)$.

## V , §3. Exercises

**1.** In this exercise, we will just compute the rank of some matrices. Nothing special here, so I'll not write all the steps involved.

**(a)** $A = \begin{pmatrix} 2 & 1 & 3 \\ 7 & 2 & 0 \end{pmatrix}$ and $\mathrm{Rank}(A) = 2$

**(b)** $A = \begin{pmatrix} -1 & 2 & -2 \\ 3 & 4 & -5 \end{pmatrix}$ and $\mathrm{Rank}(A) = 2$

**(c)** $A = \begin{pmatrix} 1 & 2 & 7 \\ 2 & 4 & -1 \end{pmatrix}$ and $\mathrm{Rank}(A) = 2$

**(d)** $A = \begin{pmatrix} 1 & 2 & -3 \\ -1 & -2 & 3 \\ 4 & 8 & -12 \\ 0 & 0 & 0 \end{pmatrix}$ and $\mathrm{Rank}(A) = 1$

**(e)** $A = \begin{pmatrix} 2 & 0 \\ 0 & -5 \end{pmatrix}$ and $\mathrm{Rank}(A) = 2$

**2.** Suppose $A$ is an $m \times n$ matrix and $B$ is an $n \times p$ matrix over some field $K$. Consider the corresponding linear maps $L_A : K^n \to K^m$ and $L_B : K^p \to K^n$. Now, the matrix $AB$ is the matrix of the linear map $L_A \circ L_B : K^p \to K^m$. Also, we have

$$\begin{aligned} \text{Rank}(AB) &= \dim \text{im} L_A \circ L_B \\ \text{Rank}(A) &= \dim \text{im} L_A \\ \text{Rank}(B) &= \dim \text{im} L_B \end{aligned}$$

Now, $\text{im} L_A \circ L_B \subseteq \text{im} L_A$, which means that $\dim \text{im} L_A \circ L_B \leq \dim \text{im} L_A$, which basically is saying that $\text{Rank}(AB) \leq \text{Rank}(A)$. Now, by Rank-Nullity theorem, we have

$$\begin{aligned} &\dim \text{Ker} L_B + \dim \text{im} L_B = p \\ &\dim \text{Ker} L_A \circ L_B + \dim \text{im} L_A \circ L_B = p \end{aligned}$$

Also, observe that $\text{Ker} L_B \subseteq \text{Ker} L_A \circ L_B$, which means that $\dim \text{Ker} L_B \leq \dim \text{Ker} L_A \circ L_B$, and so we conclude that

$$\dim \text{im} L_B \geq \dim \text{im} L_A \circ L_B$$

which is basically saying that $\text{Rank}(AB) \leq \text{Rank}(B)$. So, it can be concluded that

$$\text{Rank}(AB) \leq \min(\text{Rank}(A), \text{Rank}(B))$$

**3.** Let $A$ be an upper triangular square matrix with none of the diagonal elements 0, i.e

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ 0 & 0 & 0 & \dots & a_{4n} \\ 0 & 0 & 0 & \dots & a_{5n} \end{pmatrix}$$

with $a_{ii} \neq 0$. Then, it is very easy to see that the rows of $A$ are linearly independent, and hence $\text{Rank}(A) = n$.

**4. (d)** The given system of equations is

$$\begin{aligned} x + y + z &= 0 \\ x - y &= 0 \\ y + z &= 0 \end{aligned}$$

We can write this in matrix form as

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = O$$

Now, we know that

$$3 = \dim \text{Space of solutions} + \text{Rank}(A)$$

which means that the dimension of the space of solutions is 0. So, the only solution to this system is $(0, 0, 0)$.

**5. (c)** The given system of equations is:

$$\begin{aligned} 2x - 3y + z &= 0 \\ x + y - z &= 0 \\ 3x + 4y &= 0 \\ 5x + y + z &= 0 \end{aligned}$$

which in matrix form can be written as:

$$\begin{pmatrix} 2 & -3 & 1 \\ 1 & 1 & -1 \\ 3 & 4 & 0 \\ 5 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = O$$

Again, we know that

$$3 = \dim \text{Space of solutions} + \text{Rank}(A)$$

which means that the dimension of the space of solutions is 0, which means that the only solution of this system is $(0, 0, 0)$.

**6.** We need to find the dimension of the space of solutions of

$$X \cdot A = P \cdot A$$

which can also be written as

$$(X - P) \cdot A = 0$$

Note that, the dimension of the solution space of the above equation is equal to the dimension of the space of solutions of the equation

$$X \cdot A = 0$$

We can interpret $A$ as a row matrix, and thus $\text{Rank}(A) = 1$. So, the dimension of the space of solutions is $n - 1$.

## Row Operations and Invertible Matrices

This is a section inspired from the book "Algebra" by Michael Artin. Let $A$ be an $n \times p$ matrix. Then, there are three fundamental "row" operations related to matrices:

(1) Add a multiple of the $j^{\text{th}}$ row to the $i^{\text{th}}$ row of $A$, where $i \neq j$.
(2) Switch the $i^{\text{th}}$ and $j^{\text{th}}$ rows of $A$.
(3) Multiply the $i^{\text{th}}$ row of $A$ by a scalar $c$.

It turns out that these three operations can be carried out by multiplication with suitable 'elementary matrices'. Corresponding to every elementary row operation, there are three types of elementary matrices. Let $e_{ij}$ be an $n \times n$ matrix, all of whose entries are zero except the $ij^{\text{th}}$ entry.

(1) The first type of elementary matrix is given by the formula

$$E_1 = I_n + ce_{ij}$$

where $c$ is any scalar, and $i \neq j$. If $A$ is any $n \times p$ matrix, then the matrix $E_1A$ adds $c$ times the $j^{\text{th}}$ row to the $i^{\text{th}}$ row of $A$. Note that $E_1$ is an $n \times n$ square matrix.

(2) The second type of elementary matrix is given by the formula

$$E_2 = I_n + e_{ij} + e_{ji} - e_{ii} - e_{jj}$$

If $A$ is an $n \times p$ matrix, then the effect of $E_2A$ will be to switch the $i^{\text{th}}$ and $j^{\text{th}}$ rows of $A$. Again, $E_2$ is a $n \times n$ square matrix.

(3) The third type of elementary matrix is given by formula

$$E_3 = I_n + (c-1)e_{ii}$$

and the effect of $E_3A$ is to multiply the $i^{\text{th}}$ row of $A$ by a scalar $c$.

All the three elementary matrices are invertible, which is not hard to see. If we apply a sequence of elementary row operations $E_1$, $E_2$, ..., $E_k$ to $A$, then the resultant matrix is given by

$$E_k ... E_2 E_1 A$$

It is also not hard to see that the product of elementary matrices is also invertible.

One advantage of using row reduction is to solve linear equations. Suppose we have a system of $m$ equations in $n$ unknowns, then we can write the system as

$$AX = B$$

where $A$ is an $m \times n$ matrix, and $B$ is an $m \times 1$ matrix. If we simultaneously do row-reduction on $A$ and $B$, and obtain matrices $A'$ and $B'$, then the key fact is that the solutions of the system $A'X = B'$ will exactly be the same as the solutions of the original system. Thus, row-reduction can help to easily solve linear equations.

Using row-reductions, given any $m \times n$ matrix, we can reduce to a form known as the *row echelon form*. It is basically an $m \times n$ matrix, which has the following properties:

(1) The first non-zero entry of every row is 1. These entries are called the 'pivots'.
(2) The first non-zero entry of the $(i+1)^{\text{th}}$ row is to the right of the $i^{\text{th}}$ row.
(3) All entries above a pivot are zero.

Now, if $A$ is a square matrix, then we can reduce it using elementary row operations, to an echelon form. Also, if $M$ is a square echelon matrix, it is not hard to see that either it is the identity matrix, or its bottom row is zero. So, $A$ is invertible if and only if, it can be reduced to the identity matrix by a sequence of row operations, and that is the case if and only if it is a product of elementary matrices.

Now, if $A$ is invertible, then a sequence of row operations reduce it to $I_n$, i.e

$$E_k ... E_2 E_1 A = I$$

and we have

$$E_k ... E_2 E_1 I = A^{-1}$$

Hence, applying the same operations to the identity matrix will give us the inverse of $A$.

## Determinants

We already know how to define determinants for $2 \times 2$ and $3 \times 3$ matrices. We can inductively define determinants for an $n \times n$ matrix. We can view determinants as a multilinear function of the columns of a square matrix. We will see from the next theorem that determinants can be characterised by three properties. We will not give a complete proof of the theorem.

**Theorem 0.10.** There exists a unique function $D : K^n \times K^n \times ... \times K^n \to K$ which has the following properties:

(1) $D$ is multinlinear
(2) $D$ is *alternating*, i.e

$$D(A_1, ..., A_j, A_j, ..., A_n) = 0$$

which means that if two consecutive columns are equal, then $D$ is zero.
(3) $D(E_1, E_2, ..., E_n) = 0$, where $E_1, E_2, ..., E_n$ is the standard orthonormal basis of $K^n$. We can also write it is $D(I) = 0$, where $I$ is the identity square matrix.

*Proof:* We will not prove uniqueness here. We will only prove the existence, and we will do so by induction.

Let us define

$$D(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} D(A_{1j})$$

where $A_{1j}$ is the $(n-1) \times (n-1)$ matrix obtained by removing the $1^{\text{st}}$ row and the $j^{\text{th}}$ column of $A$. We will show that this definition satisfies the three properties.

Let's interpret $D$ as a function of columns of $A$. Now, consider the term

$$(-1)^{1+j} a_{1j} D(A_{1j})$$

and consider the $k^{\text{th}}$ column of $A$. If $j \neq k$, then the term $a_{1j}$ is independent of the column of $A$, and by induction hypothesis $D(A_{1j})$ is linear in the $k^{\text{th}}$ column of $A_{1j}$. If $j = k$, then $a_{1j}$ is linear on the $k^{\text{th}}$ column of $A$, and $D(A_{1j})$ is independent of this column. So, every term is linear on the column, and so is $D$.

The second and third properties have simple proofs. I will not include them.

Now, we will prove some interesting properties of determinants. In the below discussion, $A$ is a square $n \times n$ matrix.

**Theorem 0.11.** If consecutive columns of $A$ are interchanged, then the determinant changes sign.

*Proof:* Suppose columns $A^j$ and $A^{j+1}$ are interchanged. Then,

$$D(A^1, ..., A^j, A^{j+1}, ..., A^n) + D(A^1, ..., A^{j+1}, A^j, ..., A^n) = D(A^1, ..., A^j + A^{j+1}, A^{j+1} + A^j, ..., A^n)$$

and the last term is zero because two successive columns are equal. So the assertion follows.

**Theorem 0.12.** If two columns of $A$ are equal, then the determinant is 0.

*Proof:* Suppose the determinant is of the form

$$D(A) = D(A^1, A^2, ..., A^i, ..., A^j, ..., A^n)$$

where $A^i = A^j$. We keep on switching $A^i$ with $A^{i+1}$, until we reach a situation where two adjacent columns are equal. Since switching consecutive columns only changes the sign of the determinant, it follows that the determinant must be 0.

**Theorem 0.13.** If two columns of $A$ are interchanged, then the determinant changes sign.

*Proof:* The proof is very similar to **Theorem 0.11**, and we just use **Theorem 0.12**.

**Theorem 0.14.** Adding a scalar multiple of one column to another column doesn't change the determinant.

*Proof:* Let the required determinant be given by
$$D(A^1, ..., A^i + tA^j, ..., A^j, ..., A^n)$$
where $i \neq j$, and $t \in K$. Expanding this by linearity and taking out the constant $t$, we get the required result.

## VI, §3. Exercises

**3.** The determinant of a diagonal matrix is the product of its diagonal entries.

**6.** (i) In general, the determinant of an upper-triangular or lower-triangular matrix is the product of its diagonal elements.

## Determinants and Linear Independence

We will now prove a theorem that will tell us how to know if a set of vectors in $K^n$ is linearly independent or not:

**Theorem 0.15.** Suppose $A^1, ..., A^n$ are column vectors in the space $K^n$. If $A^1, ..., A^n$ are linearly dependent, then
$$D(A^1, ..., A^n) = 0$$

*Proof:* Suppose $A^1, ..., A^n$ are linearly dependent. So, there are numbers $a_1, ..., a_n$ in $K$, not all zero, such that
$$a_1 A^1 + ... + a_n A^n = O$$
Without loss of generality, suppose $a_1 \neq 0$. Then, we can write $A^1$ as a linear combination of $A^2, ..., A^n$, so let
$$A^1 = x_2 A^2 + ... + x^n A^n$$
Now, we see that
$$\begin{aligned} D(A^1, ..., A^n) &= D(x_2 A^2 + ... + x_n A^n, ..., A^n) \\ &= x^2 D(A^2, ..., A^n) + x_3 D(A^3, ..., A^n) + ... + x^n D(A^n, ..., A^n) \\ &= 0 \end{aligned}$$
because all the determinants on the right hand side have two equal columns. So, the vectors must be linearly dependent.

As a corollary to this, we can say that vectors $A^1, ..., A^n$ are linearly independent if the determinant of these vectors is non-zero. Later, we will prove that this is actually an if and only if condition.

At this point, we will prove another important theorem about invertibility of square matrices:

**Theorem 0.16.** Suppose $A$ is an $n \times n$ matrix over some field $K$. Then, $A$ is invertible if and only if the columns of $A$ are linearly independent.

*Proof:* First, suppose that the columns of $A$ are linearly indepedent. So, this means that there is a linear map $T : K^n \to K^n$ such that
$$\begin{aligned} T(A^1) &= E_1 \\ T(A^2) &= E_2 \\ ... &= ... \\ T(A^n) &= E_n \end{aligned}$$

where $\{E_1, E_2, ..., E_n\}$ is the standard orthonormal basis of $K^n$. Consider the matrix corresponding to $T$, and denote it by $M_T$. Then, we have

$$M_T A^i = E_i$$

but this means that

$$M_T A = I$$

and so $A$ is invertible.

Conversely, suppose $A$ is invertible. Then, the linear map $L_A$ associated with $A$ is also invertible. This means that $\mathrm{Ker} L_A = \{O\}$. Also, observe that if $X \in K^n$ such that $X = (x_1, x_2, ..., x_n)$, then

$$L_A(X) = x_1 A^1 + x_2 A^2 + ... + x_n A^n$$

and since $\mathrm{Ker} L_A = \{O\}$, it follows that $\{A^1, ..., A^n\}$ are linearly independent. So the proof is complete.

Next, we will prove a theorem about *column equivalent* matrices. We say that two matrices $A$ and $B$(over some field $K$) are *column equivalent* if $B$ can be obtained from $A$ by a finite sequence of column operations.

**Theorem 0.17.** Suppose $A$ and $B$ are $n \times n$ column equivalent matrices. Then,

$$\mathrm{Rank}(A) = \mathrm{Rank}(B),$$

$A$ is invertible if and only if $B$ is invertible, and $\det(A) = 0$ if and only if $\det(B) = 0$.

*Proof:* Let $A^1, ..., A^n$ be the columns of $A$. If we interchange columns, then the space generated by the columns still remains the same. If we multiply a column by a (non-zero) constant, then again the space generated by the columns still remains the same. Finally, adding a scalar multiple of one column to another doesn't change the space generated by the columns. So, it follows that the column ranks of $A$ and $B$ are equal, and thus

$$\mathrm{Rank}(A) = \mathrm{Rank}(B)$$

Next, observe that column operations only change the sign of the determinant or scale it by a non-zero constant. So, $\det(A) = 0$ if and only if $\det(B) = 0$.

Finally, if $A$ is invertible, then

$$\mathrm{Rank}(A) = n$$

(because $A$ is invertible if and only if its columns are linearly independent) which means that $\mathrm{Rank}(B) = n$, and hence $B$ is also invertible. Similarly if $B$ is invertible then $A$ is invertible because the rank doesn't change. Hence, the claim follows.

**Theorem 0.18.** Let $A$ be an $n \times n$ matrix. Then, $A$ is column equivalent to a triangular matrix

$$B = \begin{pmatrix} b_{11} & 0 & 0 & ... & 0 \\ b_{21} & b_{22} & 0 & ... & 0 \\ b_{31} & b_{32} & b_{33} & ... & 0 \\ ... & ... & ... & ... & ... \\ b_{n1} & b_{n2} & b_{n3} & ... & b_{nn} \end{pmatrix}$$

*Proof:* The proof is by induction. If the first row of the matrix $A$ is zero, then inductively do column operations on the $(n-1) \times (n-1)$ submatrix to get the required triangular matrix. If some element of the first row is non-zero, we can assume that the first element of the first row is non-zero(by switching columns). Then, make all elements of the first row except the first zero by adding a suitable scalar multiple of the first column to the other columns. Then, inductively do column operations on the $(n-1)\times(n-1)$ submatrix. The required triangular matrix is thus obtained. Now, let's prove the most important theorem of this section:

**Theorem 0.19.** Suppose $A = (A^1, ..., A^n)$ is an $n \times n$ matrix. Then, the following conditions are equivalent:

- $A$ is invertible
- $A^1, ..., A^n$ are linearly independent
- $\det(A) \neq 0$

*Proof:* We have already shown the equivalence of the first two conditions. Now, by the previous theorem(**Theorem 0.18**), we can assume that $A$ is a triangular matrix. We have already shown that if $\det(A) \neq 0$ then the columns of $A$ are linearly independent. Now, if the columns of $A$ are linearly independent, then none of the diagonal elements is zero. Also, for a triangular matrix, determinant is the product of diagonal elements. Hence, $\det(A) \neq 0$. So the proof is complete.

## Permutations

Let's start with some examples.

**Example 1.** We wish to express the permutation

$$\begin{bmatrix}1 & 2 & 3\\3 & 1 & 2\end{bmatrix}$$

as a product of transpositions. It is not difficult to obtain this product:

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

**Example 3.** Lets express

$$\begin{bmatrix}1 & 2 & 3 & 4\\2 & 3 & 4 & 1\end{bmatrix}$$

as a product of transpositions. Again, this is not difficult, and we obtain

$$\begin{bmatrix}1 & 2 & 3 & 4\\2 & 3 & 4 & 1\end{bmatrix} = \begin{bmatrix}1 & 2 & 3 & 4\\4 & 2 & 3 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 3 & 4\\3 & 2 & 1 & 4\end{bmatrix}\begin{bmatrix}1 & 2 & 3 & 4\\2 & 1 & 3 & 4\end{bmatrix}$$

It can actually be proven that every permutation can be represented as a product of transpositions. Lets prove that.

**Theorem 0.20.** Let $\sigma$ be a permutation of $\{1, ..., n\}$. Then, $\sigma$ can be expressed as a product of transpositions.

*Proof:* We can prove this by induction. If $n = 1$ then there is nothing to prove. Now, suppose $\sigma$ is a permutation of $\{1, 2, ..., n\}$. Let $\sigma(n) = k$. Consider the transposition $\tau$ such that $\tau(n) = k$ and $\tau(k) = n$. So, the permutation $\tau\sigma$ fixes $n$, and so it is a permutation of $\{1, 2, ..., n-1\}$ that keeps $n$ fixed. By induction hypothesis, we can write

$$\tau\sigma = \tau_1\tau_2...\tau_s$$

where each $\tau_i$ is a transposition, and thus

$$\sigma = \tau^{-1}\tau_1...\tau_s = \tau\tau_1...\tau_s$$

and the claim follows.

We will denote $\{1, ..., n\}$ by the symbol $[n]$. We now prove an important theorem.

**Theorem 0.21.** To every permutation $\sigma$ of $[n]$ we can assign a sign $\epsilon(\sigma)$, which is $-1$ or 1, and which satisfies the following properties:

- If $\tau$ is a transposition, then $\epsilon(\tau) = -1$
- If $\sigma$ and $\sigma'$ are permutations then
$$\epsilon(\sigma\sigma') = \epsilon(\sigma)\epsilon(\sigma')$$

Proof: Let $E^1, E^2, ..., E^n$ be the standard orthonormal basis of $K^n$. Let $\sigma$ be any permutation of $[n]$. We define
$$\epsilon(\sigma) = \frac{\det(E^{\sigma(1)}, E^{\sigma(2)}, ..., E^{\sigma(n)})}{\det(E^1, ..., E^n)}$$
which we can also write as the equation
$$\det(E^{\sigma(1)}, E^{\sigma(2)}, ..., E^{\sigma(n)}) = \epsilon(\sigma)\det(E^1, ..., E^n)$$
It is then clear that if $\tau$ is a transposition, then $\epsilon(\tau) = -1$.

Next, let $\sigma$ and $\sigma'$ be any two permutations. Then,
$$\det(E^{\sigma\sigma'(1)}, E^{\sigma\sigma'(2)}, ..., E^{\sigma\sigma'(n)} = \epsilon(\sigma\sigma')\det(E^1, ..., E^n)$$
But, we can also write
$$\begin{aligned}\det(E^{\sigma\sigma'(1)}, E^{\sigma\sigma'(2)}, ..., E^{\sigma\sigma'(n)} &= \epsilon(\sigma')\det(E^{\sigma(1)}, E^{\sigma(2)}, ..., E^{\sigma(n)}) \\ &= \epsilon(\sigma')\epsilon(\sigma)\det(E^1, ..., E^n)\end{aligned}$$
and the equality $\epsilon(\sigma\sigma') = \epsilon(\sigma)\epsilon(\sigma')$ immediately follows. **Note:** Using this theorem and the fact that any permutation can be written as a product of transpositions, we can easily compute the sign of a given permutation. Permutations with sign 1 are called *even*, while those with sign $-1$ are called *odd*.

## VI, §6. Exercises

**1.** In this exercise we will determine the sign of some permutations:

(b). $\epsilon\left(\begin{bmatrix}1 & 2 & 3\\ 3 & 1 & 2\end{bmatrix}\right) = (-1)^2 = 1$

(f). $\epsilon\left(\begin{bmatrix}1 & 2 & 3 & 4\\ 3 & 2 & 4 & 1\end{bmatrix}\right) = (-1)^2 = 1$

(g). $\epsilon\left(\begin{bmatrix}1 & 2 & 3 & 4\\ 4 & 2 & 1 & 3\end{bmatrix}\right) = (-1)^2 = 1$

**3.** We will show that the number of even permutations of $[n]$ is equal to the number of odd permutations. We will establish a bijection between the even and the odd permutations.

First, let $\tau$ be any transposition. Then, consider the map given by $\sigma \mapsto \tau\sigma$. Suppose there are permutations $\sigma_1$ and $\sigma_2$ such that
$$\tau\sigma_1 = \tau\sigma_2$$
Then it is clear that $\sigma_1 = \sigma_2$, and so the map is injective. To show that the map is surjective, observe that for any permutation $\sigma$, we have
$$\tau(\tau^{-1}\sigma) = \sigma.$$
Finally, since $\tau$ is an odd permutation, for any even permutation $\sigma$, we have that $\tau\sigma$ is an odd permutation. So, this is a bijection between even and odd permutations. Hence the claim follows.

We will now use our knowledge of permutations to prove that any alternating multilinear function $D$ (multinear on columns) such that $D(E_1, ..., E_n) = 1$ is uniquely determined by these three properties.

First, let's try to prove it for the simpler $2 \times 2$ case. Let

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be any $2 \times 2$ matrix, and observe that

$$A^1 = aE^1 + cE^2$$
$$A^2 = bE^1 + dE^2$$

where $E^1$ and $E^2$ are the standard orthonormal column vectors. Then,

$$\begin{aligned} D(A^1, A^2) &= D(aE^1 + cE^2, bE^1 + dE^2) \\ &= aD(E^1, bE^1 + dE^2) + cD(E^2, bE^1 + dE^2) \\ &= abD(E^1, E^1) + adD(E^1, E^2) + cbD(E^2, E^1) + cdD(E^2, E^2) \\ &= ad - bc \end{aligned}$$

Note that we used some additional properties of determinants which can be proven using the three initial properties.

There was no mention of permutations here. But, let us see why this is ultimately related to permutations. We will prove it as a general theorem for $n \times n$ matrices. Before doing so, let's make a quick remark. Let $A$ be an $n \times n$ matrix with entries $a_{ij}$. Then, we can write the columns of $A$ as

$$A^1 = a_{11}E^1 + a_{21}E^2 + ...a_{n1}E^n = \sum_{j_1=1}^{n} a_{j_1 1}E^{j_1}$$

$$A^2 = a_{12}E^1 + a_{22}E^2 + ...a_{n2}E^n = \sum_{j_2=1}^{n} a_{j_2 2}E^{j_1}$$

.....

$$A^n = a_{1n}E^1 + a_{2n}E^2 + ...a_{nn}E^n = \sum_{j_n=1}^{n} a_{j_n n}E^{j_n}$$

In the following theorem, we will replace $E^1, ..., E^n$ be any general column vectors $X^1, ..., X^n$.

**Theorem 0.22.** Suppose $B$ is any $n \times n$ matrix with entries $b_{ij}$. Let $X^1, ..., X^n$ be any column vectors. Define the columns

$$\begin{aligned} A^1 &= b_{11}X^1 + b_{21}X^2 + ... + b_{n1}X^n \\ A^2 &= b_{12}X^1 + b_{22}X^2 + ... + b_{n2}X^n \\ &\quad ... \\ A^n &= b_{1n}X^1 + b_{2n}X^2 + ... + b_{nn}X^n \end{aligned}$$

Then,

$$D(A^1, ..., A^n) = \sum_{\sigma} \epsilon(\sigma) b_{\sigma(1)1} b_{\sigma(2)2} ... b_{\sigma(n)n} D(X^1, ..., X^n)$$

where the sum is taken over all permutations $\sigma$ of $[n]$.

*Proof:* By the the definition of $A^i$, we can write

$$A^i = \sum_{j_i=1}^{n} b_{j_i i} X^{j_i}$$

Then, we use multilinearity of $D$ to get

$$\begin{aligned} D(A^1, ..., A^n) &= D\left(\sum_{j_1=1}^{n} b_{j_1 1} X^{j_1}, ..., \sum_{j_n=1}^{n} b_{j_n n} X^{j_n}\right) \\ &= \sum_{j_1=1}^{n} b_{j_1 1} D\left(X^{j_i}, ..., \sum_{j_n=1}^{n} b_{j_n n} X^{j_n}\right) \\ &= ... \\ &= \sum_{j_i=1}^{n} \sum_{j_2=1}^{n} ... \sum_{j_n=1}^{n} b_{j_1 1} b_{j_2 2} ... b_{j_n n} D(X^{j_1}, ..., X^{j_n}) \end{aligned}$$

Now, in the above sum, $j_1, ..., j_n$ can take any values from 1 to $n$, but we can only focus on the cases where the $j_i$s are distinct. So, the sum becomes

$$\begin{aligned} D(A^1, ..., A^n) &= \sum_{\sigma} b_{\sigma(1)1} b_{\sigma(2)2} ... b_{\sigma(n)n} D(X^{\sigma(1)}, ..., X^{\sigma(n)}) \\ &= \sum_{\sigma} \epsilon(\sigma) b_{\sigma(1)1} b_{\sigma(2)2} ... b_{\sigma(n)n} D(X^1, ..., X^n) \end{aligned}$$

and so the formula is proven. Note that the only thing we used was multilinearity, and the fact that $D$ is alternating.

We can replace $X^1, ..., X^n$ by the vectors $E^1, ..., E^n$ to obtain that for any $n \times n$ matrix $A$, we have

$$\det(A) = \sum_{\sigma} \epsilon(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} ... a_{\sigma(n)n}$$

and hence the determinant is uniquely determined by these properties.

Let's now prove some more useful facts about determinants using this formula.

**Theorem 0.23.** Suppose $A$ and $B$ are $n \times n$ matrices. Then,

$$\det(AB) = \det(A)\det(B)$$

*Proof:* Suppose $C = AB$. Then, the $i^{\text{th}}$ column of $C$ is given by

$$C^i = b_{1i} A^1 + b_{2i} A^2 + ... + b_{ni} A^n$$

So, by **Theorem 0.22**, we have

$$\det(C) = \sum_{\sigma} \epsilon(\sigma) b_{\sigma(1)1} b_{\sigma(2)2} ... b_{\sigma(n)n} D(A^1, ..., A^n) = \det(A)\det(B)$$

and so the claim follows.

**Theorem 0.24.** For any square matrix $A$,

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

*Proof:* For any permutation $\sigma$, consider the product

$$a_{\sigma(1)1} ... a_{\sigma(n)n}$$

Suppose $\sigma(1) = k$, so we can write

$$a_{\sigma(1)1} = a_{k\sigma^{-1}(k)}$$

and so we can write

$$a_{\sigma(1)1}...a_{\sigma(n)n} = a_{1\sigma^{-1}(1)}...a_{n\sigma^{-1}(n)}$$

and so

$$\det(A) = \sum_{\sigma} \epsilon(\sigma^{-1}) a_{1\sigma^{-1}(1)}...a_{n\sigma^{-1}(n)}$$

As $\sigma$ ranges over all permutations, so does $\sigma^{-1}$. So, the second hand side of the last equation is the determinant of $A^y$. So, the claim follows.

Now, let us prove a formula for the inverse of a non-singular matrix.

**Theorem 0.25.** Let $A$ be an $n \times n$ matrix such that $D(A) \neq 0$. Then, the inverse of $A$ is given by a matrix $B$ such that

$$B_{ij} = \frac{D(A^1, ..., E^j, ..., A^n)}{D(A)}$$

where the vector $E^j$ occurs at the $i^{th}$ column in the numerator.

*Proof:* To find the inverse of $A$, we need to find a matrix $B$ such that $AB = I$, or we can write

$$E^j = B_{1j}A^1 + ... + B_{nj}A^n$$

This is a system of $n$ linear equations in $n$ unknowns. The solution can be obtained by using Cramer's rule as

$$x_{ij} = \frac{D(A^1, ..., E^j, ..., A^n)}{D(A)}$$

and so the theorem is proved.

This shows that $B$ is the right-inverse of $A$. Now we will show that $B$ is also the left inverse. Since $D(A) = D(A^t)$, we can repeat the above process to obtain a matrix $Y$ such that

$$A^t Y = I$$

and taking the transpose on both sides, we get

$$Y^t A = I$$

So,

$$BA = (Y^t A)(BA) = Y^t (AB) A = Y^t A = I$$

and hence $B$ is also the left-inverse of $A$.

Define $A_{ij}$ to be the matrix obtained after removing the $i^{\text{th}}$ row and the $j^{\text{th}}$ column of $A$. So, observe that

$$B_{ij} = \frac{(-1)^{i+j} \det(A_{ji})}{\det(A)}$$

and so we can say that the inverse of $A$ is the transpose of the matrix of cofactors of $A$ divided by $\det(A)$.

## The Binet-Cauchy Formula

In this section, we will see a generalisation of the product rule for determinants.

## 1. VI , §8. Exercises

**1.** As an exercise, we will compute the inverse of a matrix using the above formula.

The matrix is

$$\begin{pmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 1 & 1 \end{pmatrix}$$

and the inverse by the above formula is

$$\frac{-1}{20}\begin{pmatrix} 4 & 1 & -7 \\ -4 & -6 & 2 \\ -7 & 2 & 6 \end{pmatrix}$$

## Internal and External Direct Sums

Suppose $V$ is a vector space over a field $K$. Let $V_1$ and $V_2$ be subspaces of $V$ which satisfy the following properties:

(1) $V_1 + V_2 = V$
(2) $V_1 \cap V_2 = \{O\}$

Then we say that $V$ is a *direct sum* of $V_1$ and $V_2$, and we write

$$V = V_1 \oplus V_2$$

This is also called the *internal direct sum*, because $V_1$ and $V_2$ are subspaces of $V$.
Now, suppose $V_1$ and $V_2$ are abstract vector spaces over a field $K$. Now define

$$V = V_1 \times V_2$$

Now, define

$$\widetilde{V_1} = V_1 \times \{O_W\}$$

and also

$$\widetilde{V_2} = \{O_V\} \times \{V_2\}$$

Then it is not hard to see that

$$V = \widetilde{V_1} + \widetilde{V_2}$$

Also, the nice thing about $\widetilde{V_1}$ and $\widetilde{V_2}$ is that $V_1 \cong \widetilde{V_1}$ and $V_2 \cong \widetilde{V_2}$. This kind of direct sum is called an *external direct sum*, because $V_1$ and $V_2$ are no more subspaces of $V$. But due to the isomorphism, any external direct sum can be interpreted as an internal direct sum.

## Quotient Spaces

Let $V$ be a vector space over some field $K$, and let $W$ be a subspace of $V$. We define a relation $\backsim$ on $V$ as follows: $v \backsim w$ if $v - w \in W$. It is then easy to see that $\backsim$ is actually an equivalence relation. The set of all equivalence classes under this relation is called the *quotient space*, and is denoted by ${}^{V}\!/_{W}$. For a vector $v$, we denote its equivalence class by the symbol $[v]$. Also, the equality

$$[v] = v + W$$

is true, where

$$v + W := \{u + v : u \in W\}$$

Now, we will define addition and scalar multiplication of equivalence classes as follows:

$$[v_1] + [v_2] = [v_1 + v_2]$$
$$\alpha[v_1] = [\alpha v_1]$$

where $\alpha \in K$. This shows that ${}^{V}/_{K}$ is itself a vector space, whose zero element is $[O] = W$. We now define a map $\pi : V \to {}^{V}/_{W}$ as

$$\pi(v) = [v]$$

It is then easily seen that $\text{Ker}\pi = W$, and $\text{im}\pi = {}^{V}/_{W}$. And by Rank-Nullity theorem, we get

$$\dim {}^{V}/_{W} = \dim V - \dim W$$

This gives us an important fact: given any subspace $W$ of $V$, there is a linear map from $V$ to some vector space whose kernel is $W$. Let's do an example of quotient spaces.

**Example**. Let $V = \mathbb{R}^2$. Let $L_0$ be the subspace of $V$ given by

$$L_0 = \{(x, 0) : x \in \mathbb{R}\}$$

We wish to find ${}^{V}/_{L_0}$. The elements of this quotient space are the translates of $L_0$, which are basically elements of the form $v + L_0$, where $v \in V$. So, the elements are $L_\alpha$, where $\alpha$ is any real number, and

$$L_\alpha = \{(x, \alpha) : x \in \mathbb{R}\}$$

which means that the quotient space consists of all lines in $\mathbb{R}^2$ parallel to the line $y = 0$.

We will now prove what's called the *first isomorphism theorem*.

**Theorem 1.1.** Let $V$ and $W$ be vector spaces over some field $K$, and let $T : V \to W$ be a linear map. Then,

$${}^{V}/_{\text{Ker}T} \cong \text{im}T$$

*Proof:* We will give an isomorphism from ${}^{V}/_{\text{Ker}T}$ to $\text{im}T$.

For $v + \text{Ker}T \in {}^{V}/_{\text{Ker}T}$, define

$$P(v + \text{Ker}T) = T(v)$$

We must ensure that this map is well-defined, in the sense that if $v_1 + \text{Ker}T = v_2 + \text{Ker}T$ then $P(v_1 + \text{Ker}T) = P(v_2 + \text{Ker}T)$, but this is easily checked. We will show that $P$ is a one-one and onto linear map.

First let's show that it is a linear map. Let $v_1 + \text{Ker}T$ and $v_2 + \text{Ker}T$ be in ${}^{V}/_{\text{Ker}T}$. Then,

$$\begin{aligned}
P(v_1 + \text{Ker}T + v_2 + \text{Ker}T) &= P(v_1 + v_2 + \text{Ker}T) \\
&= T(v_1 + v_2) \\
&= T(v_1) + T(v_2) \\
&= P(v_1 + \text{Ker}T) + T(v_2 + \text{Ker}T)
\end{aligned}$$

and if $\alpha \in K$, then

$$\begin{aligned}
P(\alpha(v_1 + \text{Ker}T)) &= P(\alpha v_1 + \text{Ker}T) \\
&= T(\alpha v_1) \\
&= \alpha T(v_1) \\
&= \alpha P(v_1 + \text{Ker}T)
\end{aligned}$$

To show that $P$ is one-one, we will prove that $\text{Ker}P = \{O + \text{Ker}T\}$ ($O$ is the zero vector in $V$). Suppose there is some $v + \text{Ker}T$ such that

$$P(v + \text{Ker}T) = T(v) = O$$

(observe that here $O$ is the zero element of $W$). This means that $v \in \text{Ker}T$, which means that $v + \text{Ker}T = 0 + \text{Ker}T$. So, $T$ is one-one.

Finally, showing that $P$ is onto is trivial. So, it follows that $P$ is an isomorphism, and hence

$$V/_{\text{Ker}T} \cong \text{im}T$$

## General Orthogonal Bases and The Algebraic Dual

In this section, we will first prove that any non-zero finite dimensional vector space $V$ has an orthogonal basis.

**Theorem 1.2.** Let $V$ be a finite dimensional vector space over a field $K$, equipped with a scalar product, and suppose $V \neq \{O\}$. Then, $V$ has an orthogonal basis.

*Proof:* We can prove this by induction on the dimension of $V$. If $V$ is one-dimensional, then any non-zero vector in $V$ forms an orthogonal basis. So, suppose $n = \dim V > 1$. Then, two cases are possible.

**Case 1:** For every vector $v \in V$, we have

$$\langle v, v \rangle = 0$$

and a consequence of this is that for every $v, w \in V$, we have

$$\langle v, w \rangle = 0$$

which means that every basis of $V$ is orthogonal.

**Case 2:** There is some vector $v_1 \in V$ such that $\langle v_1, v_1 \rangle \neq 0$. Let $V_1$ be the space generated by the vector $v_1$. We will show that

$$V = V_1 \oplus V_1^{\perp}$$

To show this, suppose $v \in V$, and let

$$c = \frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle}$$

and observe that $\langle v - cv_1, v_1 \rangle = 0$, which means that $v - cv_1 \in V_1^{\perp}$ and hence

$$v = (v - cv_1) + cv_1$$

Also, we must have $V_1 \cap V_1^{\perp} = \{O\}$, which proves our claim. Now, $\dim V_1^{\perp} = \dim V - 1$, and by induction hypothesis, $V_1^{\perp}$ has an orthogonal basis $\{v_2, ..., v_n\}$, and it follows that

$$\{v_1, v_2, ..., v_n\}$$

is an orthogonal basis of $V$. This proves the claim.

## V,§5. Exercises

**1.** In this exercise we will find orthogonal basis for the indicated spaces.

**(a)**. $A = (1, 1, 1)$, $B = (1, -1, 2)$ and $X \cdot Y = x_1y_1 + 2x_2y_2 + x_3y_3$. It is clear that $A$ and $B$ are linearly independent, but

$$\langle A, B\rangle \neq 0$$

Also, observe that $\langle A, A\rangle \neq 0$. Consider the vector $B'$ given by

$$B' = B - \frac{\langle B, A\rangle}{\langle A, A\rangle}A = \left(\frac{3}{4}, \frac{-5}{4}, \frac{7}{4}\right)$$

So, the set $\{A, B'\}$ is an orthogonal basis for the given space.

**2.** Consider the space $\mathbb{C}^2$ over $\mathbb{C}$. For elements $(x_1, x_2)$ and $(y_1, y_2)$ in $\mathbb{C}^2$, we define

$$\langle (x_1, x_2), (y_1, y_2)\rangle = x_1y_1 - ix_2y_1 - ix_1y_2 - 2x_2y_2$$

Consider the elements $A = (1, 1)$ and $B = (1, -1)$, which form a basis of $\mathbb{C}^2$. Also, we have

$$\langle (1, 1), (1, 1)\rangle = -1 - 2i \neq 0$$

Now, define $B'$ as

$$B' = B - \frac{\langle B, A\rangle}{\langle A, A\rangle}A = \left(\frac{8 - 6i}{5}, \frac{-2 - 6i}{5}\right)$$

So, $\{A, B'\}$ is an orthogonal basis of $\mathbb{C}^2$.

Suppose $V$ is a vector space over some field $K$. We define the space $V^*$ as

$$V^* := \{T : V \to K : T \text{ is linear}\}$$

Observe that $V^*$ is a subspace of $K^V$. We call this space the *algebraic dual* of $V$.

Suppose $\dim V = N$. We already know that $V \cong K^N$. We will prove that in this case, $V^* \cong K^N$ as well, which will prove that $V \cong V^*$.

Suppose $\{v_1, ..., v_N\}$ is a basis of $V$. Any $f \in V^*$ is completely determined by the values $f(v_1), ..., f(v_N)$. So, consider the map $\phi : V^* \to K^N$ defined as

$$\phi(f) = (f(v_1), ..., f(v_n))$$

It is then clear that $\phi$ is a one-one map. Also, it is clearly onto. Finally, it is also easy to see that $\phi$ is a linear map. Hence, $\phi$ is an isomorphism, and hence

$$V \cong V^*$$

Note that $V$ was assumed to be finite dimensional. It turns out that this is a necessary and sufficient condition for the dual space to be isomorphic to the original space.

We will now look at finite dimensional vector spaces with a *non-degenerate* scalar product. It turns out that the dual space in this case is isomorphic to something simple, as we prove in the following theorem:

**Theorem 1.3.** Suppose $V$ is a finite dimensional vector space with a non-degenerate scalar product. For any $v \in V$, define the map $L_v(w) = \langle v, w\rangle$. Then, $L_v \in V^*$. The map

$$v \mapsto L_v$$

is an isomorphism between $V$ and $V^*$.

In other words, this theorem means that any map $\phi \in V^*$ can be written in the form

$$\phi(v) = \langle v_0, v\rangle$$

for some element $v_0 \in V$.

*Proof:* First, the mapping $v \mapsto L_v$ is clearly linear by the properties of the scalar product.

Next, suppose there is some $v \in V$ such that $L_v = O$, where $O$ is the zero-map. This means that for all $w \in V$, we have

$$\langle v, w \rangle = 0$$

and since the scalar product is non-degenerate, this implies that $v = O_V$. So, the kernel of the mapping $v \mapsto L_v$ is $\{O_V\}$. By the Rank-Nullity theorem, it follows that the dimension of the image of this map is $\dim V$. Since $\dim V = \dim V*$, it follows that this map is surjective. Hence it is an isomorphism.

Let's prove one final theorem in this section:

**Theorem 1.4.** Suppose $V$ is a finite dimensional vector space, and let $W$ be a subspace. Define

$$W^{\perp} := \{\phi \in V^* : \phi(W) = 0\}$$

Then, $W^{\perp}$ is a subspace of $V^*$, and

$$\dim W + \dim W^{\perp} = \dim V$$

*Proof:* That $W^{\perp}$ is a subspace of $V^*$ is easy to see. If $W = \{O_V\}$, then $W^{\perp} = V^*$, and the claim follows. So, suppose $W \neq \{O_V\}$. Let $\{w_1, w_2, ..., w_r\}$ be a basis of $W$, and extend it to a basis of $V$, that is let

$$\{w_1, w_2, ..., w_r, w_{r+1}, ..., w_n\}$$

be a basis of $V$. Corresponding to this basis, let $\{\phi_1, \phi_2, ..., \phi_n\}$ be the dual basis. We will show that $\{\phi_{r+1}, \phi_{r+2}, ..., \phi_n\}$ is a basis of $W^{\perp}$.

First, any linear combination of $\{\phi_{r+1}, ..., \phi_n\}$ is in $W^{\perp}$. Next, if $\phi \in W^{\perp}$, then we can write

$$\phi = a_1\phi_1 + ... + a_n\phi_n$$

and observe that

$$\begin{aligned} \phi(w_1) &= a_1 = 0 \\ \phi(w_2) &= a_2 = 0 \\ &.... \\ \phi(w_r) &= a_r = 0 \end{aligned}$$

and hence $\phi$ is a linear combination of $\{\phi_{r+1}, ..., \phi_n\}$. This proves the theorem.

Observe that in the above theorem no use of the scalar product is made. But we will now use it to prove a corollary.

Let $V$ be a finite dimensional vector space with a *non-degenerate* scalar product. For a subspace $W$ of $V$, we define it's *orthogonal complement* in two ways:

$$\begin{aligned} \mathrm{perp}_V(W) &= \{v \in V : \langle v, w \rangle = 0 \ \forall w \in W\} \\ \mathrm{perp}_{V^*}(W) &= \{\phi \in V^* : \phi(W) = 0\} \end{aligned}$$

Since the scalar product is degenerate, the map $v \mapsto L_v$ gives an isomorphism between $\mathrm{perp}_V(W)$ and $\mathrm{perp}_{V^*}(W)$, where $L_v(w) = \langle v, w \rangle$ for all $w \in V$. So, as a corollary to the previous theorem, we obtain:

**Theorem 1.5.** Suppose $V$ is a finite dimensional vector space with a non-degenerate scalar product. For any subspace $W$ of $V$, we have

$$\dim V = \dim W + \dim W^{\perp}$$

Here, $W^{\perp}$ is the set of all vectors in $V$ which are orthogonal to every vector in $W$.

## V, §6. Exercises

**3.** Suppose $W$ is the subspace of $\mathbb{C}^3$ generated by the vector $(1, i, 0)$. We know that the dimension of $W^{\perp}$ is $3-1=2$. So, let's find two linearly independent vectors which are orthogonal to $(1, i, 0)$. Any such vector $(x, y, z) \in \mathbb{C}^3$ will satisfy the equation

$$x + iy = 0$$

Observe that two such vectors are $(i, -1, 1)$ and $(1, i, 0)$, both of which are linearly independent. Hence, the set $\{(i, -1, 1), (1, i, 0)\}$ is the basis of $W^{\perp}$.

**4.** Suppose $V$ is an $n$ dimensional vector space over $K$, and consider $\phi$, a non-zero functional (a map in $V^*$). By Rank-Nullity theorem, we have

$$n = \dim \operatorname{Ker}(\phi) + \dim \operatorname{Im}(\phi)$$

and since the image is a non-zero subspace of $K$, it must be $K$. Hence,

$$\dim \operatorname{Ker}(\phi) = n - 1.$$

## Space of Operators

In this section we will study something known as the *space of operators*. Let $V$ and $W$ be finite dimensional vector spaces over a field $K$. The set of all linear maps (or operators) from $V$ to $W$ is denoted by $\operatorname{Hom}(V, W)$. This is a vector space over $K$ as we proven before.

We define $\operatorname{End}(V) = \operatorname{Hom}(V, W)$. Observe that not only is $\operatorname{End}(V)$ a vector space, but it is equipped with another operation, which is the composition of linear maps. Let $T \in \operatorname{End}(V)$. As a notation, we write

$$K[T] := \{c_0 I + c_1 T + ... + c_n T^n : c_0, ..., c_n \in K\}$$

Here, $I$ is the identity map in $\operatorname{End}(V)$. Note that if $g \in K[T]$ and $f \in K[T]$, then we have

$$f \circ g = g \circ f$$

that is composition is commutative in $K[T]$. Let's again come back to the space $\operatorname{Hom}(V, W)$. We will find the dimension of this space.

**Theorem 1.6.** If $V$ and $W$ are finite-dimensional vector spaces over $K$, then

$$\dim \operatorname{Hom}(V, W) = \dim(V) \dim(W)$$

*Proof:* Let $\{v_1, ..., v_n\}$ be a basis of $V$, and let $w_1, ..., w_m$ be a basis of $W$. Here, the key fact that we will use is that any $T \in \operatorname{Hom}(V, W)$ is completely determined by the values $T(v_1), ..., T(v_n)$.

For $1 \le i \le n$ and $1 \le j \le m$, define $T_{ij} \in \operatorname{Hom}(V, W)$ by the formula

$$T_{ij}(v_k) = \delta_{ik} w_j$$

for any $1 \le k \le n$. In simpler words, the map $T_{ij}$ sends the vector $v_i$ to $w_j$, and sends all other vectors to the zero vector. We will show that the set $\{T_{ij} : 1 \le i \le n, 1 \le j \le m\}$ is a basis of $\operatorname{Hom}(V, W)$, which will prove the claim.

First, let $\lambda_{ij}$ be scalars in $K$ such that

$$f = \sum_{\substack{1\leq i\leq n\\1\leq j\leq m}} \lambda_{ij}T_{ij} = O$$

Here, $O$ is the zero map. Let's compute $f(v_1)$. We have

$$\begin{aligned}f(v_1) &= \sum_{\substack{1\leq i\leq n\\1\leq j\leq m}} \lambda_{ij}T_{ij}(v_1)\\&= \sum_{\substack{1\leq i\leq n\\1\leq j\leq m}} \lambda_{ij}\delta_{i1}w_j\\&= \sum_{j=1}^{m}\lambda_{1j}w_j\\&= O_W\end{aligned}$$

and since $w_1, ..., w_m$ are linearly independent, it follows that $\lambda 1j = 0$ for each $1 \leq j \leq m$. Similarly, by computing $f(v_k)$, we can show that $\lambda_{kj} = 0$ for each $1 \leq j \leq m$. This proves that $\lambda_{ij} = 0$ for all $i$ and $j$, and hence the set $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ is linearly independent.

Now, let $T : V \to W$ be a linear map which takes the following values on the basis elements of $V$:

$$\begin{aligned}T(v_1) &= a_{11}w_1 + a_{12}w_2 + ... + a_{1m}w_m\\T(v_2) &= a_{21}w_1 + a_{22}w_2 + ... + a_{2m}w_m\\&...\\T(v_n) &= a_{n1}w_1 + a_{n2}w_2 + ... + a_{nm}w_m\end{aligned}$$

and this can be written as

$$\begin{aligned}T(v_1) &= a_{11}T_{11}(v_1) + a_{12}T_{12}(v_1) + ... + a_{1m}T_{1m}(v_1)\\T(v_2) &= a_{21}T_{21}(v_2) + a_{22}T_{22}(v_2) + ... + a_{2m}T_{2m}(v_2)\\&...\\T(v_n) &= a_{n1}T_{n1}(v_n) + a_{n2}T_{n2}(v_n) + ... + a_{nm}T_{nm}(v_n)\end{aligned}$$

and so we can write the map $T$ as

$$T = \sum_{\substack{1\leq i\leq n\\1\leq j\leq m}} a_{ij}T_{ij}$$

which proves that the set $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ is a basis of $\mathrm{Hom}(V, W)$. So, it directly follows that

$$\dim \mathrm{Hom}(V, W) = \dim(V) \dim(W)$$

In particular, the above formula tells us that

$$\dim \mathrm{End}(V) = (\dim(V))^2$$

if $V$ is finite dimensional.

**Polynomials.** Let $V$ be a vector space over some field $K$, and consider the objects $\text{End}(V)$ and $K[x]$. As a convention, we define

$$\deg O = \infty$$

where $O \in K[x]$ is the zero polynomial. We already know an important fact about the ring $K[x]$: it has Euclidean division. Let's now prove the following theorem:

**Theorem 1.7.** Suppose $V$ is finite dimensional, with $\dim V = N$. Let $T \in \text{End}(V)$. Then, there is a non-zero polynomial $f \in K[x]$ of degree less than or equal to $N^2$ for which the map $f(T)$ is the zero map.

*Proof:* Consider the following maps in $\text{End}(V)$

$$I, T, T^2, T^3, ..., T^{N^2}$$

where $I$ is the identity mapping. We know that these must be linearly dependent, because $\dim \text{End}(V) = N^2$, and so there are scalars $c_0, ..., c_{N^2}$ not all zero such that

$$c_0 I + c_1 T + c_2 T^2 + ... + c_{N^2} T^{N^2} = O$$

where $O$ is the zero map. Consequently, the polynomial $f \in K[x]$ defined by

$$f(x) = c_0 + c_1 x + ... + c_{N^2} x^{N^2}$$

is non-zero, and the map $f(T)$ is the zero map. This proves the claim.

Let's make a couple of definitions. Let $T \in \text{End}(V)$. We say that $T$ is *regular/invertible* if there is some map $S \in \text{End}(V)$ such that $S \circ T = I = T \circ S$. It $T$ is not regular, then it is called *singular*. We also define the *general linear* group $\text{GL}(V)$ to be the set

$$\text{GL}(V) := \{T \in \text{End}(V) : T \text{ is invertible}\}$$

(Its called a group because it forms a group under composition). If $V$ is finite dimensional, then $T \in \text{End}(V)$ is invertible if and only if $\text{Ker}T = \{O\}$. This is easily seen by the Rank-Nullity theorem.

Another definition. For $T \in \text{End}(V)$, we define the *annihilator* of $T$ to be the set

$$\text{An}(T) = \{f \in K[x] : f(T) = O\}$$

where $O \in \text{End}(V)$ is the zero map. Let's prove a relatively simple theorem.

**Theorem 1.8.** Let $T \in \text{End}(V)$. Then:

(1) $f, g \in \text{An}(T) \implies f + g \in \text{An}(T)$
(2) $f \in \text{An}(T)$ and $g \in K[x]$ implies that $fg \in \text{An}(T)$ and $gf \in \text{An}(T)$.

*Proof:* For the first assertion, suppose $f, g \in \text{An}(T)$, and let

$$f = c_0 + ... + c_n x^n$$
$$g = b_0 + ... + b_m x^m$$

and wlog let $m > n$. Then, observe that

$$(f + g)(T) = (c_0 + b_0)I + ... + (c_n + b_n)T^n + ...c_m T^m$$

and for any element $v \in V$, we have

$$\begin{aligned}(f + g)(T)(v) &= (c_0 + b_0)I(v) + ... + (c_n + b_n)T^n(v) + ... + c_m T^m(v) \\ &= c_0 I(v) + ... + c_m T^m(v) + b_0 I(v) + ... + b_n T^n(v) \\ &= O + O \\ &= O\end{aligned}$$

and hence $f + g \in \mathrm{An}(V)$. In a similar way, we can prove the second assertion.

Now suppose $V$ is finite dimensional. Then, by **Theorem 1.3**, there is some non-zero polynomial $f \in K[x]$ such that $f \in \mathrm{An}(T)$, for any $T \in \mathrm{End}(V)$. Also, if this $f$ satisfies

$$f(x) = c_0 + ... + c_r x^r$$

then by **Theorem 1.4**, we see that $c_r^{-1} f \in \mathrm{An}(T)$, and observe that $c_r^{-1} f$ is a monic polynomial. With this in mind, for a given $T$, we define its *minimal polynomial* to be the monic polynomial of least degree which annihilates $T$. By Euclidean Division in $K[x]$ combined with **Theorem 1.4**, it follows that this minimal polynomial is unique.

**Examples**. Following are some examples of linear maps and their minimal polynomials:

**(a)**The minimal polynomial of the zero map is $x$.

**(b)**The minimal polynomial of the identity mapping $I$ is $x - 1$. The minimal polynomial of the map $\lambda I$ is $x - \lambda$.

**(c)** Let's see an example where the minimal polynomial is $x^2 + 1$. Suppose $V = \mathbb{R}^2$, and let $\{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$. Define a map $T \in \mathrm{End}(V)$ as

$$T(e_1) = e_2$$
$$T(e_2) = -e_1$$

It is then easy to see that no monic polynomial of degree one can be the minimal polyomial. Also, observe that

$$T^2 + I = O$$

and hence its minimal polynomial is $x^2 + 1$.

Let's make some more definitions. We say that $T \in \mathrm{End}(V)$ is *nilpotent* there is some $n \geq 1$ such that $T^n = O$. Let's do prove an easy theorem:

**Theorem 1.9.** If $T$ is nilpotent, then both $T - I$ and $T + I$ are invertible.

*Proof:* The conclusion follows from the simple observation that

$$I = (T - I)(-I - T - ... - T^{n-1})$$
$$I = (-T - I)(-I + T - T^2 ... + (-1)^n T^{n-1})$$

Let's do another important fact about invertible operators:

**Theorem 1.10.** Suppose $\dim V < \infty$, and suppose $T \in \mathrm{End}(V)$. Let $P_T$ be the minimal polynomial of $T$. Then, $T$ is invertible if and only if $P_T(0) \neq 0$

*Proof:* First, suppose that $P_T(0) \neq 0$. Then, let

$$P_T(x) = c_0 + c_1 x + ... + x^n$$

where $n \geq 1$. So, we have

$$c_0 I + c_1 T + ... + T^n = O$$

which implies that

$$I = T\left(\frac{-c_1}{c_0} I - ... - \frac{T^{n-1}}{c_0}\right)$$

which implies that $T$ is invertible.

We will prove the other direction by proving the contrapositive. Suppose $P_T(0) = 0$. Then, let

$$P_T(x) = c_1 x + ... + x^n$$

which means that

$$c_1 T + ... + T^n = O$$

where $n \geq 1$. If $n = 1$, this means that

$$T = O$$

and hence $T$ is not invertible. If $n > 1$, then we have

$$T(c_1 I + ... + T^{n-1}) = O$$

and since $P_T$ is the minimal polynomial for $T$, it must be true that

$$c_1 I + ... + T^{n-1}$$

is not the zero map. So, there is some non-zero $w \in V$ such that $(c_1 I + ... + T^{n-1})(w) \neq O_V$, which means that $T(w) = O_W$, and hence $\text{Ker}(T)$ is non-zero, which means that $T$ is not invertible. This completes the proof.

So, to check whether a given operator is invertible or not, we just need to check the constant of its minimal polynomial.

Let's make some more definitions. Suppose $V$ is finite dimensional, and let $T \in \text{End}(V)$. $\lambda \in K$ is called an *eigenvalue* of $T$ if $T - \lambda I$ is singular(not invertible). This is equivalent to saying that there is some non-zero $v \in V$ such that

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

.

Let's prove an important theorem about eigenvalues of an operator:

**Theorem 1.11.** Let $V$ be a finite dimensional vector space, and let $T$ be an operator on $V$. Then, $\lambda \in K$ is an eigenvalue of $T$ if and only if $\lambda$ is the root of $P_T$.

*Proof:* We might as well assume that $T$ is the non-zero operator. Suppose $\lambda$ is an eigenvalue of $T$. We have that

$$P_T(x) = (x - \lambda)q(x) + c$$

where $c \in K$. Let $v \in V$ be the corresponding non-zero eigenvector for this eigenvalue. Then, we have

$$\begin{aligned} P_T(T)(v) = O_V &= (T - \lambda I) \circ q(T)(v) + cI(v) \\ &= q(T) \circ (T - \lambda I)(v) + cI(v) \\ &= O_V + cv \end{aligned}$$

which implies that $cv = O_V$, and since $v \neq O_V$, it implies that $c = 0$. So, $\lambda$ is a root of $P_T$.

Conversely, suppose $\lambda$ is a root of $P_T$. Then,

$$P_T(x) = (x - \lambda)q(x)$$

for some non-zero polynomial $q$. This means that $(T - \lambda I)q(T)$ is the zero-map. But, $q(T)$ cannot be the zero-map, and so there exists some non-zero $v \in V$ such that $(T - \lambda I)(v) = O_V$, which means that $\lambda$ is an eigenvalue of $T$. This completes the proof.

Now let us prove another important theorem:

**Theorem 1.12.** Let $T \in \text{End}(V)$, and let $\lambda$ be an eigenvalue of $T$. Let $V_\lambda$ be the set of all eigenvectors of $T$ corresponding to this eigenvalue. Then, $V_\lambda$ is a subspace of $V$.

Proof: It is clear that $O$ in $V_\lambda$. Also, suppose $v_1$, $v_2$ are in $V_\lambda$. Then,

$$T(v_1 + v_2) = \lambda v_1 + \lambda v_2 = \lambda(v_1 + v_2)$$

and hence $v_1 + v_2 \in V_\lambda$. It can be similarly shown that $cv_1 \in V_\lambda$ for any $c \in K$. So, the proof is complete.

This space is called the *eigenspace* corresponding to $\lambda$.

Let us now prove another theorem:

**Theorem 1.13.** Eigenvectors corresponding to distinct eigenvalues are linearly independent.

*Proof:* Suppose $T \in \mathrm{End}(V)$ and let $a_1, ..., a_n$ be distinct eigenvalues of $T$, with corresponding eigenvectors $v_1, ..., v_n$. We prove the claim by induction. Clearly, a single vector is linearly independent. So, let $v_1, ..., v_r$ be a subset of $v_1, ..., v_n$, and suppose all sets of $r - 1$ vectors are linearly independent. Now, suppose there are scalars $c_1, ..., c_r$ such that

$$c_1 v_1 + ... + c_r v_r = O$$

Appplying $T$ on both sides, we get

$$a_1 c_1 v_1 + ... + a_r c_r v_r = O$$

and this means that

$$c_2(a_2 - a_1)v_2 + ... + c_r(a_r - a_1)v_r = O$$

and since this is a set of $r - 1$ vectors, these are linearly independent. Since $a_i - a_1$ is not zero for $i \neq 1$, we see that

$$c_2 = ... = c_r = 0$$

and hence $c_1 = 0$ (because $v_1 \neq O$). So, by induction, it follows that $v_1, ..., v_n$ are linearly independent.

Now, we will prove a theorem which will lead to a concept called *diagonalisation.*

**Theorem 1.14.** Suppose $T \in \mathrm{End}(V)$. The following hold:

(1) $T \in \mathrm{GL}(V)$ if and only if 0 is not an eigenvalue of $T$.

(2) For any $S \in \mathrm{GL}(V)$, we have

$$(STS^{-1})^n = ST^nS^{-1}$$

(3) For any $f \in K[x]$, we have

$$f(STS^{-1}) = Sf(T)S^{-1}$$

(4) The operators $T$ and $STS^{-1}$ have the same eigenvalues.

*Proof:* For (1), we know that $T$ is invertible if and only if $P_T(0) \neq 0$. Also, $P_T(0) \neq 0$ if and only if 0 is not an eigenvalue of $T$ (because eigenvalues are roots of $P_T$). Hence the claim follows.

For (2), we prove it using induction. The case $n = 1$ is trivial. Then, observe that

$$\begin{aligned}(STS^{-1})^{n+1} &= (STS^{-1})^n(STS^{-1}) \\ &= (ST^nS^{-1})(STS^{-1}) \\ &= ST^{n+1}S^{-1}\end{aligned}$$

and so we are done by induction.

For (3), suppose $f \in K[x]$ is given by

$$f(x) = c_0 + c_1x_1 + ... + c_nx^n$$

Then, we have

$$\begin{aligned} f(STS^{-1}) &= c_0SIS^{-1} + c_1STS^{-1} + c_2ST^2S^{-1} + ... + c_nST^nS^{-1} \\ &= S(c_0IS^{-1} + c_1TS^{-1} + c_2T^2S^{-1} + ... + c_nT^nS^{-1}) \\ &= S(c_0I + c_1T + c_2T^2 + ... + c_nT^n)S^{-1} \\ &= Sf(T)S^{-1} \end{aligned}$$

Note that we first used left distributivity of composition, then we used right distributivitiy. For (4), we will prove something stronger: we will show that the minimal polynomials of $STS^{-1}$ and $T$ are equal.

Let $g$ be the minimal polynomial of $STS^{-1}$, and let $f$ be the minimal polynomial of $T$.

Then, observe that

$$g(STS^{-1}) = O = Sg(T)S^{-1}$$

where $O$ is the zero map. Since $S \in \mathrm{GL}(V)$, it follows that $g(T)$ is the zero map, and hence $g \in \mathrm{An}(T)$. So, $f|g$. Similarly, observe that

$$f(STS^{-1}) = Sf(T)S^{-1} = O$$

and hence $f \in \mathrm{An}(STS^{-1})$, so that $g|f$. Since $f$ and $g$ are both monic polynomials in $K[x]$ that divide each other, it follows that $f = g$. So, the minimal polynomials are equal, and hence the eigenvalues are also equal.

Respecting the above theorem, we define a *diagonalisable* operator. $T \in \mathrm{End}(V)$ is said to be *diagonalisable* if there is a basis of $V$ which consists of eigenvalues.

Let's now prove a key theorem about diagonalisable operators:

**Theorem 1.15.** $T \in \mathrm{End}(V)$ is diagonalisable if and only if all roots of $P_T$ are in the field $K$ and each root has multiplicity 1.

*Proof:* First, suppose $T$ is diagonalisable. Then, there is a basis $\{v_1, ..., v_n\}$ of $V$ that consists of eigenvectors of $T$. Suppose

$$\begin{aligned} T(v_1) &= \lambda_1 v_1 \\ T(v_2) &= \lambda_2 v_1 \\ &... \\ T(v_n) &= \lambda_n v_n \end{aligned}$$

Consider

$$F(x) = (x - \lambda_1)...(x - \lambda_r)$$

where $1 \leq r \leq n$, and each $\lambda_i$ is distinct (this is done to ensure that each $\lambda_i$ has multiplicity 1 in $F$). Now, observe that

$$F(T) = O$$

and hence $F$ annihilates $T$, so $P_T|F$. Also, we know that each $\lambda_i$ is an eigenvalue, and hence $(x - \lambda_i)|P_T$ for each $i$. But, observe that if $i \neq j$, then $(x - \lambda_i)$ and $(x - \lambda_j)$ are coprime, and hence it follows that $(x - \lambda_1)...(x - \lambda_r)|P_T$, which means that $F|P_T$. Finally, both $F$ and $P_T$ are monic polynomials which divide each other, and hence $F = P_T$. So, each root of $P_T$ is in $K$ and has multiplicity 1.

Conversely, suppose that $P = (x - \lambda_1)...(x - \lambda_r)$, where each $\lambda_i$ is distinct, and $1 \leq r \leq n$. Let $W_1, ..., W_r$ be the eigenspaces corresponding to the eigenvalues $\lambda_1, ..., \lambda_r$

respectively. We will show that $V$ is the direct sum of $W_1, ..., W_r$, and hence it will show that $T$ is diagonalisable.

It is clear that $W_i \cap W_j = \{O\}$ if $i \neq j$. So, all we need to prove is that every $v \in V$ can be written in the form

$$v = w_1 + ... + w_r$$

, where $w_i \in W_i$.

First, suppose $w \in W_i$, for any $i$. Observe that

$$T(T(w)) = T(\lambda_i w) = \lambda_i T(w)$$

and hence $T(w) \in W_i$. We say that each space $W_i$ is $T$*-invariant.*

Consider the polynomials

$$Q_i(x) = \frac{P(x)}{x - \lambda_i}$$

for each $1 \leq i \leq r$. Then, the polynomials $Q_1, ..., Q_r$ are relatively prime. So, there are polynomials $G_1, ..., G_r$ such that

$$Q_1 G_1 + ... + Q_r G_r = 1$$

where 1 is the constant polynomial.

First, we will show that for any $v \in V$, $Q_i(T)(v) \in W_i$. Since $P_T$ is the minimal polynomial, we have

$$P_T(v) = O$$

for all $v \in V$. Hence

$$(T - \lambda_i I) \circ (T - \lambda_1 I) \circ ... \circ (T - \lambda_{i-1} I) \circ (T - \lambda_{i+1} I) \circ ... \circ (T - \lambda_r I)(v) = O$$

for all $v \in V$, which means that

$$(T - \lambda_i I)(Q_i(T)(v)) = O$$

for all $v \in V$, and hence $Q_i(T)(v) \in W_i$.

Finally, observe that

$$Q_1 G_1(T)(v) + ... + Q_r G_r(T)(v) = I(v) = v$$

for all $v \in V$. This proves that we can write any $v \in V$ in the form $v = w_1 + ... + w_r$, and hence $V$ is the direct sum of $W_1, ..., W_r$. Hence, $T$ is diagonalisable, and the proof is complete.

Let us now prove another important theorem regarding upper-triangular maps (or matrices):

**Theorem 1.16.** Suppose $V$ is a finite dimensional vector space over $K$, and let $T \in \text{End}(V)$ such that there is basis $B$ for which $M_B(T)$ is upper triangular, i.e

$$M_B(T) = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ 0 & a_{22} & .. & a_{2n} \\ .. & .. & .. & .. \\ 0 & 0 & 0 & a_{nn} \end{pmatrix}$$

Then, $a_{ii}$ is an eigenvalue of $T$, for $1 \leq i \leq n$.

*Proof:* Let the basis $B$ be $\{v_1, ..., v_n\}$. Then, the following hold:

$$\begin{aligned} T(v_1) &= a_{11}v_1 \\ T(v_2) &= a_{12}v_1 + a_{22}v_2 \\ &\dots \\ T(v_n) &= a_{1n}v_1 + ... + a_{nn}v_n \end{aligned}$$

A simple calculation shows that

$$\begin{aligned} &(T - a_{11}I)(v_1) = O \\ &(T - a_{11}I) \circ (T - a_{22}I)(v_2) = O \\ &\dots \\ &(T - a_{11}I) \circ (T - a_{22}I) \circ ... \circ (T - a_{nn}I)(v_n) = O \end{aligned}$$

First, it is clear that $a_{11}$ is an eigenvalue. Now suppose that $a_{11}, ..., a_{k-1k-1}$ are eigenvalues. If $a_{kk}$ is any one of $a_{11}, ..., a_{k-1k-1}$ then $a_{kk}$ is also an eigenvalue. So, suppose $a_{kk}$ is not one of those values. Observe that

$$(T - a_{k-1k-1}I) \circ ... \circ (T - a_{11}I)(v_k) = d_1v_1 + ... + d_kv_k$$

where $d_1, ..., d_k$ are in $K$. A simple calculation shows that

$$d_k = (a_{kk} - a_{k-1k-1})...(a_{kk} - a_{11})$$

and hence $d_k \neq 0$. Hence, $(T - a_{k-1k-1}I) \circ ... \circ (T - a_{11}I)(v_k) \neq O$. But, observe that

$$(T - a_{kk}I) \circ (T - a_{k-1k-1}I) \circ ... \circ (T - a_{11}I)(v_k) = O$$

and hence it follows that $a_{kk}$ is also an eigenvalue of $T$. Hence, all the diagonal elements are eigenvalues of $T$, and the proof is complete.

Next, let's prove that upper triangulizable and lower triangulizable are essentially the same thing, and hence we can just say a *triangulizable operator*:

**Theorem 1.17.** $T \in \text{End}(V)$ is upper triangulizable if and only if it is lower triangulizable.

*Proof*: Suppose $T$ is upper triangulizable. Then, there is some basis $\{v_1, ..., v_n\}$ such that

$$\begin{aligned} T(v_1) &= a_{11}v_1 \\ T(v_1) &= a_{12}v_1 + a_{22}v_2 \\ &\dots \\ T(v_n) &= a_{1n}v_1 + ... + a_{nn}v_n \end{aligned}$$

Consider the basis $\{v_n, ..., v_1\}$. The matrix of $T$ with respect to this basis is then:

$$\begin{pmatrix} a_{nn} & 0 & \dots & 0 \\ a_{n-1n} & a_{n-1n-1} & \dots & 0 \\ a_{n-2n} & a_{n-2n-1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{1n-1} & \dots & a_{11} \end{pmatrix}$$

which is a lower triangular matrix. Hence, $T$ is lower triangulizable. Similarly, the other direction may be proved, and hence the claim follows.

## VIII , §1. Exercises

**1.** Suppose $a \in K$ and $a \neq 0$. It is easy to see that the only eigenvalue of the matrix

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

is $\lambda = 1$, and any eigenvector corresponding to this eigenvalue is of the form $(\alpha, 0)$, and a basis for this eigenspace is $\{(1,0)\}$. Evidently, this is a one-dimensional space.

**2.** Consider the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

and again the only eigenvalue is $\lambda = 2$, and every vector is an eigenvector. So, a basis of the eigenspace is $\{(1,0),(0,1)\}$.

**3.** Suppose $A$ is a diagonal matrix given by

$$A = \begin{pmatrix} a_{11} & 0 & ... & 0 \\ 0 & a_{22} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & a_{nn} \end{pmatrix}$$

Suppose $E_1, ..., E_n$ are the standard orthonormal basis. Then observe that

$$\begin{aligned} A(E_1) &= a_{11}E_1 \\ A(E_2) &= a_{22}E_2 \\ &... \\ A(E_n) &= a_{nn}E_n \end{aligned}$$

and hence the dimension of the space generated by the eigenvectors of $A$ is $n$ (caution: this does not mean that every vector is an eigenvector). All diagonal elements are the eigenvalues, and their corresponding eigenvectors are these vectors.

**5.** **(a)** Suppose $\theta \in \mathbb{R}$ and consider the matrix

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

First, suppose $\cos\theta \neq 1$. Consider the vector $v$ given by

$$v = \left( \frac{\sin\theta}{1-\cos\theta}, 1 \right)$$

It is not hard to see that $Av = v$. Hence, in this case, $A$ has an eigenvector with a real eigenvalue. If $\cos\theta = 1$, then the matrix $A$ is simply

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

and the vector $v = (1,0)$ is an eigenvector of $A$ with corresponding eigenvalue 1.

**6.** Let $R(\theta)$ be a rotation matrix in $\mathbb{R}^2$. The characteristic polynomial of this matrix is $t^2 - 2t\cos\theta + 1$. This polynomial has real roots if and only if $\sin\theta = 0$, which is equivalent to saying that $R(\theta) = \pm I$.

## The Characteristic Polynomial

We have already shown before that if $T \in \text{End}(V)$, then $\lambda$ is an eigenvalue of $T$ if and only if $T - \lambda I$ is not invertible. If $T$ is interpreted as a matrix, then we define the *characteristic polynomial* of $T$ to be the polynomial

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

where $I$ is the identity matrix. Let's prove an important theorem:

**Theorem 1.18.** $\lambda \in K$ is an eigenvalue of $T$ if and only if $\lambda$ is a root of the characteristic polynomial of $T$. Here $T$ is a matrix.

*Proof:* Suppose $\lambda \in K$ is an eigenvalue of $T$. Then, we know that the matrix $T - \lambda I$ is not invertible, and hence $\det(T - \lambda I) = 0$. Conversely, suppose $\lambda \in K$ is a root of the characteristic polynomial. Then, we have

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

and hence $T - \lambda I$ is not invertible, and hence $\lambda$ is an eigenvalue of $T$.

The characteristic polynomial gives us a good computational tool to calculate the eigenvalues of a map.

Let's also prove a relationship between determinants of similar matrices:

**Theorem 1.19.** Suppose $A$ and $B$ are two $n \times n$ matrices, such that $B$ is invertible. Then, the characteristic polynomials of $A$ and $B^{-1}AB$ are equal.

*Proof:* We have that

$$\begin{aligned} \det(A - tI) &= \det(B^{-1}(A - tI)B) \\ &= \det(B^{-1}AB - tI) \end{aligned}$$

and this proves our theorem.

## VIII , §2. Exercises

**1.** **(a)** The characteristic polynomial of $A$ is $(x - a_1)(x - a_2)...(x - a_n)$.
**(b)** The eigenvalues of $A$ are $a_1, ..., a_n$.

**2.** If $A$ is a triangular matrix, the answer is still as it was in **1**.

**9.** Suppose $V$ is an $n$ dimensional vector space and suppose that the characteristic polynomial of a map $A \in \text{End}(V)$ has $n$ distinct roots. Then, it implies that $A$ has $n$ distinct eigenvalues, and hence $A$ is diagonalisable. This means that there is some basis of $V$ consisting of eigenvalues of $A$.

**10.** We will prove something more general: suppose $V$ is an $n$ dimensional vector space with a non-degenerate scalar product. Then, for any map $A \in \text{End}(V)$, the eigenvalues of $A$ and ${}^tA$ are equal.

*Proof:* Suppose $\lambda$ is an eigenvalue of $A$ and let $v_1 \in V$ be the corresponding eigenvector. Then, for every $w \in V$, we have

$$\langle Av, w\rangle = \langle \lambda v, w\rangle = \langle v, {}^tAw\rangle$$

which implies that

$$\langle v, \lambda w - {}^tAw\rangle = 0$$

Now, if $\lambda$ is not an eigenvalue, then the map $\lambda I - {}^tA$ is invertible, and hence an isomorphism. So, the expression $\lambda w - {}^tAw$ runs through all vectors in $V$ as $w$ runs through all vectors in $V$. But since the scalar product is non-degenerate, it implies that $v = O$, which is a contradiction. So, $\lambda$ must be an eigenvalue of ${}^tA$.

Similarly, we can show the other way: if $\lambda$ is an eigenvalue of ${}^tA$, then it is also an eigenvalue of $A$. Hence, this completes the proof.

This theorem implies that the eigenvalues of a matrix and its transpose are the same.

**15.** Suppose $V$ is a finite dimensional vector space over $K$. Let $A$ and $B$ be linear operators. We will show that $AB$ and $BA$ have the same eigenvalues. Suppose $\lambda \neq 0$ is an eigenvalue of $AB$. Then, there is some non-zero vector $v$ such that

$$ABv = \lambda v$$

Now, observe that $B(v)$ is a non-zero vector. Also, we have

$$BABv = B\lambda v = \lambda Bv$$

which implies that $\lambda$ is also an eigenvalue of $BA$. We can show the other direction similarly. Hence, the eigenvalues of $AB$ and $BA$ are the same.

Now we deal with the case when $\lambda = 0$. So, there is some non-zero $v$ such that

$$ABv = O$$

Now, if $Bv \neq O$, then we are done, because in that case,

$$BABv = O$$

and hence $Bv$ is the required eigenvector. If $Bv = O$, then there are two cases: first, if $\mathrm{Ker}(A) \neq \{O\}$, then there is some $v \neq O$ such that $Av = O$, and hence

$$BAv = O$$

so that $v$ is the required eigenvector. If $\mathrm{Ker}(A) = \{O\}$, then $A$ is an isomorphism. Hence, there is some $w \in V$ such that $A(w) = v$. Hence, in that case, we have

$$BAw = Bv = O$$

and hence $w$ is the required non-zero eigenvector. So, in all cases, $AB$ and $BA$ have the same eigenvalues.

## Semi-Simple Operators

Let's define what a semi-simple operator is. Let $V$ be a vector space over some field $K$. $T \in \mathrm{End}(V)$ is called *semisimple* if for every $T$-invariant subspace of $V$, there is a $T$-invariant complement. $T$ is called a *simple* operator if the only $T$-invariant subspaces of $V$ are $\{O\}$ and $V$.

Let's look at two examples:

**Example 1:** $T : V \to V$, such that $T(e_1) = e_1$ and $T(e_2) = e_1 + e_2$. We will show that $T$ is not semi-simple.

Suppose $T$ is semi-simple. Consider the subspace spanned by $e_1$. Since $e_1$ is an eigenvector, this subspace is $T$-invariant. So, there is a $T$-invariant complement as well, and call it $W$. Since $V$ is two-dimensional, $\dim W = 1$, and let $\{e\}$ be a basis of $W$. Since $W$ is $T$-invariant, we must have

$$T(e) = ce$$

for some $c \in K$.

Now, it is clear that $c \neq 1$, because $T$ is not the identity operator. Now, suppose $e_2 = ae_1 + be$, where $a, b \in K$. We then have

$$T(e_2) = ae_1 + cbe = (1+a)e_1 + be$$

which implies that $b(c-1) = 0$, and since $b \neq 0$, we have $c = 1$, which is a contradiction. Hence, $T$ is not semi-simple.

**Example 2:** Suppose $V$ is a two dimensional vector space over $\mathbb{R}$, and let $T$ be given by $T(e_1) = -e_2$ and $T(e_2) = e_1$. It is clear that the minimal polynomial of $T$ is $x^2+1$, and hence it is not diagonalisable. But, we will show that $T$ is simple, and hence semi-simple. This is a good example to show the fact that if the ground field is not algebraically closed, then semi-simpleness and diagonalisable are not the same thing.

Suppose $W$ is a $T$-invariant subspace of dimension 1. Let $\{e\}$ be a basis of $W$. Since $W$ is $T$-invariant, it follows that $T(e) = ce$, for some $c \in \mathbb{R}$, which means that $c$ is an eigenvalue, which contradicts the fact that $x^2+1$ is the minimal polynomial. Hence, $V$ cannot have any $T$-invariant subspaces of dimension 1, and hence it is simple.

Now, lets prove another useful theorem:

**Theorem 1.20.** Suppose $T$ is an operator on $V$ such that $T$ is diagonalisable. Then, $T$ is semi-simple.

*Proof:* We prove this by reverse induction on the degree of the $T$-invariant subspace. Suppose $W$ is a $T$-invariant subspace of $V$ of dimension $n$. Then, it follows that $W = V$, and hence the $T$-invariant complement is $\{O\}$.

Now, suppose $\dim W = r < n$, and suppose it is true that all $T$-invariant subspaces of $V$ of dimension > r have a $T$-invariant complement. Let $\{w_1, ..., w_r\}$ be a basis of $W$. Since $T$ is diagonalisable, there is a basis $\{v_1, ..., v_n\}$ of $V$ consisting of eigennvectors of $T$. Now, in this basis, there is atleast one vector $v_k$ such that the set $\{w_1, ..., w_r, v_k\}$ is linearly independent. Without loss of generality suppose $v_k = v_1$, and the set $\{w_1, ..., w_r, v_1\}$ is linearly independent. Define $W_1 = \langle\{w_1, ..., w_r, v_1\}\rangle$, and hence $\dim W_1 = r+1$. Also, since $v_1$ is an eigenvector, it follows that $W_1$ is a $T$-invariant subspace of $V$. Hence by our hypothesis, it has a $T$-invariant complement, say $W_2$. Let $\{u_{r+2}, ..., u_n\}$ be a basis of $W_2$.

Now, define

$$\widetilde{W} = \langle\{v_1, u_{r+2}, ..., u_n\}\rangle$$

Then, $\widetilde{W}$ is a $T$-invariant subspace of $V$, and observe that

$$V = W \oplus \widetilde{W}$$

and hence, $W$ has a $T$-invariant complement. So, it follows that all $T$-invariant subspaces of $V$ have a $T$-invariant complement, and our proof is complete.

Let's now prove another important theorem:

**Theorem 1.21.** $T \in \mathrm{End}(V)$ is triangulisable if and only if all the roots of the minimal polynomial are in the ground field.

*Proof:* First, suppose $T$ is triangulisable, and suppose there is a basis $\{v_1, ..., v_n\}$ of $V$ such that the following equations hold:

$$\begin{aligned}
T(v_1) &= a_{11}v_1 \\
T(v_2) &= a_{12}v_1 + a_{22}v_2 \\
&\dots \\
T(v_n) &= a_{1n}v_1 + a_{nn}v_n
\end{aligned}$$

We earlier showed that the polynomial $(x - a_{11})...(x - a_{nn})$ annihilates $T$. Hence, $P_T$ divides this polynomial, and hence it follows that all roots of $P_T$ are in the ground field.

We prove the converse by induction. If $\dim V = 1$, then the statement is trivial. So, suppose the statement holds for some $n - 1 \in \mathbb{N}$, and suppose $\dim V = n$. Suppose $T$ is an operator on $V$ for which $P_T$ has all its roots in the ground field. We know that $\deg(P_T) \geq 1$, and hence there is some root $\lambda$ of $P_T$, and this root is also an eigenvalue. Let $v_1$ be the corresponding eigenvector. Consider the subspace $W = \langle v_1 \rangle$. Since $v_1$ is an eigenvector, $W$ is a $T$-invariant subspace of $V$.

Now, consider the quotient space ${}^{V}\!/_{W}$. We define an operator $\overline{T}$ on this space as follows: for $v + W \in {}^{V}\!/_{W}$, define $\overline{T}(v + W) = T(v) + W$, i.e the equivalence class of $v$ is sent to the equivalence class of $T(v)$. We will now verify that this is a well defined operator. So, suppose $x, y \in V$ such that $x - y \in W$. Then, $T(x - y) \in W$, because $W$ is $T$-invariant, and this means that $T(x) - T(y) \in W$. So, this means that $T(x) + W = T(y) + W$, and hence the map $\overline{T}$ is well defined.

Now, we will show that $P_T(\overline{T})$ is the zero map in the quotient space. So, suppose $v + W \in {}^{V}\!/_{W}$, and suppose

$$P_T(x) = c_0 + c_1x + c_2x^2 + ... + x^k$$

for some $k \in \mathbb{N}$. Then, we have

$$P_T(\overline{T}) = c_0\overline{I} + c_1\overline{T} + c_2\overline{T}^2 + ... + \overline{T}^n$$

where $\overline{I}$ is the identity operator in the quotient space.

Now,

$$\begin{aligned}
P_T(\overline{T})(v + W) &= c_0(v + W) + c_1\overline{T}(v + W) + c_2\overline{T}^2(v + W) + ... + \overline{T}^n(v + W) \\
&= (c_0v + W) + (c_1T(v) + W) + (c_2T^2v + W) + ... + (T^n(v) + W) \\
&= (c_0v + c_1T(v) + ... + T^n(v)) + W \\
&= P_T(T)(v) + W \\
&= O + W
\end{aligned}$$

and hence $P_T$ annihilates $\overline{T}$. So, it follows that $P_{\overline{T}}$ divides $P_T$, and hence all roots of $P_{\overline{T}}$ are in the ground field. By induction hypotheses, it follows that $\overline{T}$ is triangulisable. So, let $\{v_2 + W, ..., v_n + W\}$ be the required basis of ${}^{V}\!/_{W}$ with respect to which the matrix of $\overline{T}$ is triangular.

Now, we will first show that $\{v_1, ..., v_n\}$ is a basis of $V$. So, suppose there are constants $c_i$ such that

$$c_1v_1 + ... + c_nv_n = O$$

This means that
$$c_1(v_1 + W) + ... + c_n(v_n + W) = O + W$$
in the quotient space. But, $v_1 \in W$, and hence $c_1(v_1 + W) = O + W$. So, it follows that
$$c_2(v_2 + W) + ... + c_n(v_n + W) = O + W$$
but because the elements $v_2 + W, ..., v_n + W$ are linearly independent, it follows that
$$c_1 = c_2 = ... = c_n = 0$$
and hence $\{v_1, ..., v_n\}$ is a basis of $V$.

Moreover, observe that
$$\overline{T}(v_j + W) = c_2(v_2 + W) + ... + c_j(v_j + W)$$
for $2 \leq j \leq n$ and for some constants $c_i$. Hence, we have
$$T(v_j) + W = (c_2v_2 + ... + c_jv_j) + W$$
for each $2 \leq j \leq n$, and hence
$$T(v_j) - (c_2v_2 + ... + c_jv_j) \in W$$
for each $2 \leq j \leq n$, which implies that
$$T(v_j) = c_1v_1 + ... + c_jv_j$$
for some scalar $c_1$. Hence, the matrix of $T$ with respect to the basis $\{v_1, ..., v_n\}$ is triangular, and hence $T$ is triangulisable. This completes the proof.

As a corollary to this, it can be inferred that if the ground field is algebraically closed (like $\mathbb{C}$), then every operator is triangulisable.

We now know that a diagonalisable operator is both semi-simple and triangulisable. We now show the converse, but before that we will prove a lemma:

**Lemma:** Suppose $T \in \text{End}(V)$ is semi-simple. Let $W$ be a $T$-invariant subspace of $V$. Then, the restriction of $T$ to $W$ is also semi-simple.

*Proof:* Let $W_1$ be a $T$-invariant subspace of $W$. Then, $W_1$ is also a $T$-invariant subspace of $V$. Hence, there is a $T$-invariant complement of $W_1$, and call it $W_2$. Now, consider the space $W \cap W_2$, which is also $T$-invariant. It is clear that $W_1 \cap (W \cap W_2) = \{O\}$, because $W_1 \cap W_2 = \{O\}$. Also, suppose $w \in W$. Then, we can write $w = w_1 + w_2$, for some $w_1 \in W_1$ and $w_2 \in W_2$, and hence $w_2 \in W \cap W_2$. Hence it follows that
$$W = W_1 \oplus (W \cap W_2)$$
and hence the restriction of $T$ on $W$ is also semi-simple.

Let's now move to the main theorem:

**Theorem 1.22.** Suppose $T \in \text{End}(V)$ is both triangulisable and semi-simple. Then, $T$ is diagonalisable.

*Proof:* We prove it by induction on $\dim V$. The theorem is clear if $\dim V = 1$. So, suppose $\dim V > 1$. Since $T$ is triangulisable on $V$, there exists a basis $\{v_1, ..., v_n\}$ of $V$ such that the matrix corresponding to this basis is upper triangular. Now, put $W = \langle v_1, ..., v_{n-1} \rangle$, so that $\dim W = n - 1$. It is clear that $W$ is a $T$-invariant subspace of $V$. So, there is a $T$-invariant complement of $W$, say $\widetilde{W}$. The dimension of $\widetilde{W}$ is 1, and let $\{v\}$ be a basis of $\widetilde{W}$. It is clear that $v$ is an eigenvector.

Also, the restriction of $T$ on $W$ is triangulisable, and this restriction is also semi-simple by our lemma. So, by induction hypothesis, the restriction of $T$ on $W$ is diagonalisable. So, there is some basis $\{u_1, ..., u_{n-1}\}$ of $W$ consisting of eigenvectors of $W$.

Finally, it follows that $\{u_1, ..., u_{n-1}, v\}$ is a basis of $V$ consisting of eigenvectors of $V$. Hence, $T$ is diagonalisable.

We now look at an example.

**Example 1:** Consider the map $T$ given by $T(e_1) = 0$, $T(e_2) = -e_3$ and $T(e_3) = e_2$, where the ground field is $\mathbb{R}$. We will show that $T$ is not diagonalisable, not triangulisable, but it is semi-simple.

The minimal polynomial of $T$ is $x(x^2 + 1)$, and hence $T$ is not diagonalisable, and hence not triangulisable. Now, we will prove that $T$ is semi-simple. So, suppose $W$ is a one-dimensional $T$-invariant subspace of $V$. Let $\{v_1\}$ be a basis of $W$. Then, $v_1$ must be an eigenvector, and hence the eigenvalue must be 0, so $T(v_1) = 0$. We show that $\{v_1, e_2, e_3\}$ are linearly independent. So, suppose there are $a, b, c$ in $\mathbb{R}$ such that

$$av_1 + be_2 + ce_3 = O$$

and applying $T$ to both sides, we obtain that $a = b = c = 0$, and hence $\{v_1, e_2, e_3\}$ are linearly independent. Hence, it is a basis, and the required $T$-invariant complement of $W$ is $\langle e_2, e_3\rangle$.

Observe that $\dim \mathrm{Ker}T = 1$ and hence $\dim \mathrm{Im}T = 2$. Now, suppose $W$ is a two-dimensional $T$-invariant subspace of $V$. Consider the map $T : W \to W$. Let's call this restriction map $T'$. Now, $\dim \mathrm{Ker}T'$ cannot be 2, because the dimension of the kernel of the original map was 1. We will deal with two cases: when $\dim \mathrm{Ker}T' = 1$ and when $\dim \mathrm{Ker}T' = 0$.

First, suppose $\dim \mathrm{Ker}T' = 1$. Then, $\mathrm{Ker}T' = \mathrm{Ker}T = \langle e_1\rangle$. Also, $\dim \mathrm{Im}T' = 1$, and let $\{u\}$ be the basis of this image, where $u \in W$. This means that $u$ must be an eigenvalue of $T'$, and since the only eigenvalue is 0, it follows that $T'(u) = O$, and hence $u \in \mathrm{Ker}T'$, which means that $u = ke_1$, for some scalar $k$. Hence, $\mathrm{Im}T' = \mathrm{Ker}T' = \langle e_1\rangle$. Now, consider a basis $\{e_1, v_1\}$ of $W$, where $v_1 \in W$. Then, $T(v_1) = \lambda e_1$, where $\lambda \neq 0$ (if $\lambda$ were 0, we will get two linearly independent elements of $\mathrm{Ker}T'$). We will show that $\{e_1, v_1, e_3\}$ are linearly independent. So, suppose there are scalars $a, b, c$ in $\mathbb{R}$ such that

$$ae_1 + bv_1 + ce_3 = O$$

and applying $T$ on both sides, we get

$$bT(v_1) + ce_2 = O$$

which is the same as saying

$$b\lambda e_1 + ce_2 = O$$

and hence $b = c = 0$ (because $\lambda \neq 0$). And hence, $a = b = c = 0$, and the required $T$-invariant complement of $W$ is $\langle e_3\rangle$.

Now, suppose $\dim \mathrm{Ker}T' = 0$, and hence $\mathrm{Im}T' = W$. Also, we have that $\mathrm{Im}T = \langle\{e_2, e_3\}\rangle$, and hence it follows that $W = \langle\{e_2, e_3\}\rangle$. Hence, a possible $T$-invariant complement is $\langle\{e_1\}\rangle$. Hence, $T$ is semi-simple, as we have found $T$-invariant complements in all cases.

**Example 2:** Consider the map given by $T(e_1) = e_1$, $T(e_2) = -e_3$ and $T(e_3) = e_2$. The minimal polynomial of this map is $(x-1)(x^2+1)$. Clearly, it is not diagonalisable and not triangularisable either. We will show that $T$ is semi-simple.

Suppose $W$ is a one-dimensional $T$-invariant subspace of $V$, and let its basis be $\{v_1\}$. Then, $v_1$ must be an eigenvector, and the only possible eigenvalue is 1, and hence $T(v_1) = v_1$. We will show that $\{v_1, e_2, e_3\}$ are linearly independent. So, suppose there are scalars $a, b, c$ in $\mathbb{R}$ such that

$$av_1 + be_2 + ce_3 = O$$

and applying $T$ on both sides, we get

$$av_1 + bT(e_2) + cT(e_3) = O$$

By subtracting the first equation from the second, we get that $b = c = 0$, and hence $a = 0$. So, the $T$-invariant complement of $W$ is $\{e_2, e_3\}$.

Now, it is not hard to see that $\mathrm{Im}T = V$, and hence $\dim \mathrm{Ker}T = 0$. Suppose $W$ is a two dimensional $T$-invariant subspace of $V$, and let $\{v_1, v_2\}$ be its basis. Extend this to a basis of $V$, and say the new basis if $\{v_1, v_2, v_3\}$. We will show that $T(v_3)$ is not in $\{v_1, v_2\}$

Finally, let's prove what's called the **Cayley-Hamilton** theorem.

**Theorem 1.23.** Suppose $T$ is an operator on $V$, and let $C_T$ be the characteristic polynomial of $T$. Then, $C_T(T) = O$, and hence the minimal polynomial of $T$ has degree at most $\dim V$.

*Proof:* We will prove the corresponding statement for an $n \times n$ matrix, and that will prove the theorem for the operator as well. Let $M$ be the matrix of $T$. We consider three cases. First, if $M$ is triangular, then it clear that $C_T(M) = O$. If $M$ is triangulisable, then there is an invertible matrix $S$ such that $SMS^{-1}$ is triangular. Now, the characteristic polynomial is invariant under conjugation, and hence

$$C_T(SMS^{-1}) = O = SC_T(M)S^{-1}$$

and hence $C_T(M) = O$. Finally, suppose $M$ is a general matrix. If the ground field is $K$, then let $\overline{K}$ be the corresponding algebraically closed field. Then, $M$ can also be viewed as a matrix over the field $\overline{K}$. Now, since $\overline{K}$ is algebraically closed, $M$ is triangulisable. Hence, $C_T(M) = O$, and this proves the theorem.

This establishes a fundamental fact about the minimal polynomial: its degree is at most the dimension of the vector space.

## Symmetric and Unitary Operators

First, let us study Symmetric Operators. Let $V$ be a finite dimensional vector space over a field $K$ with a non-degenerate scalar product. Let's prove the following two theorems which will justify the meaning of *transpose*:

**Theorem 1.24.** Let $A \in \mathrm{End}(V)$. Then, there exists a unique $B \in \mathrm{End}(V)$ such that

$$\langle A(v), w\rangle = \langle v, B(w)\rangle$$

for all $v, w$ in $V$. This map $B$ is also called the *transpose* of $A$, and is denoted by $A^t$.

*Proof:* First, fix $w \in V$. Consider the map

$$L(v) = \langle A(v), w\rangle$$

Then, $L \in V^*$, where $V^*$ is the dual space. Since $V$ is finite dimensional and the scalar product is non-degenerate, there exists $w'$ in $V$ such that

$$L(v) = \langle v, w'\rangle$$

Let's denote $w'$ by $B(w)$. Then, $B$ is a map from $V$ to itself. We just need to show that it is linear.

So, suppose $w_1$ and $w_2$ are in $V$. Then, for any $v \in V$,

$$\begin{aligned}\langle v, B(w_1 + w_2)\rangle &= \langle A(v), w_1 + w_2\rangle \\ &= \langle A(v), w_1\rangle + \langle A(v), w_2\rangle \\ &= \langle v, B(w_1)\rangle + \langle v, B(w_2)\rangle \\ &= \langle v, B(w_1) + B(w_2)\rangle\end{aligned}$$

which means that since the scalar product is non-degenerate, it means that $B(w_1 + w_2) = B(w_1) + B(w_2)$. Similarly, if $c \in K$, then for any $v \in V$,

$$\begin{aligned}\langle v, B(cw_1)\rangle &= \langle A(v), cw_1\rangle \\ &= c\langle A(v), w_1\rangle \\ &= c\langle v, B(w_1)\rangle \\ &= \langle v, cB(w_1)\rangle\end{aligned}$$

and again since the scalar product is non-degenerate, it follows that $B(cw_1) = cB(w_1)$, and so $B \in \mathrm{End}(V)$.

That $B$ is unique is easy to see, because the scalar product is non-degenerate.

Because the scalar product is non-degenerate, the transpose also satisfies the following properties which are not difficult to prove: (here $A$ and $B$ are in $\mathrm{End}(V)$)

(1) $(A + B)^t = A^t + B^t$
(2) $(AB)^t = B^t A^t$
(3) $(cA)^t = cA^t$
(4) $A^{tt} = A$

## VII, §1. Exercises

**5.** Suppose $V$ is a finite dimensional vector space over a field $K$ with a non-degenerate scalar product. Let $v_0, w_0$ be fixed elements of $V$. Let $A \in \mathrm{End}(V)$ given by

$$A(v) = \langle v_0, v\rangle w_0$$

Consider the map $A^t$. For any $v, w$ in $V$, we have

$$\langle A(v), w\rangle = \langle v, A^t(w)\rangle$$

Now, observe that

$$\begin{aligned}\langle A(v), w\rangle &= \langle\langle v_0, v\rangle w_0, w\rangle \\ &= \langle v_0, v\rangle\langle w_0, w\rangle \\ &= \langle v, v_0\rangle\langle w_0, w\rangle \\ &= \langle v, \langle w_0, w\rangle v_0\rangle\end{aligned}$$

and hence $A^t(v) = \langle w_0, v\rangle v_0$ for all $v \in V$.

**6.** Let $V$ be the vector space over $\mathbb{R}$ of infinitely differentiable functions vanishing outside the interval $(0, 1)$. Now, we know that the scalar product is positive definite, hence it is non-degenerate. So, there is some operator $D^t$ such that for every $f, g$ in $V$, we have

$$\langle Df, g\rangle = \langle f, D^t g\rangle$$

Observe the following integration by parts identity for such functions

$$\int_0^1 f'(x)g(x)dx = -\int_0^1 f(x)g'(x)dx = \int_0^1 f(x)(-g'(x))dx$$

and hence it follows that $D^t = -D$, which means that $D$ is anti-symmetric.

**9.** Suppose $A$, $B$ and $C$ are three symmetric matrices such that $A < B$ and $B < C$. Let $X \neq O$ be any vector in $K^n$. Then, we have

$$\begin{aligned} {}^tX(C-A)X &= {}^tX((C-B)+(B-A))X \\ &= {}^tX(C-B)X + {}^tX(B-A)X \\ &> 0 \end{aligned}$$

and hence $A < C$.

**10.** Suppose $V$ is a finite dimensional vector space over $\mathbb{R}$ with a positive definite scalar product. Suppose $V = W \oplus W^{\perp}$. Let $P$ be the projection on $W$, and suppose $W \neq \{O\}$. First, let us show that $P$ is semipositive.

Let $v \in V$ such that $v = w + \widetilde{w}$, where $w \in W$ and $\widetilde{w} \in W^{\perp}$. Then,

$$\begin{aligned} \langle Pv, v\rangle &= \langle w, v\rangle \\ &= \langle w, w + \widetilde{w}\rangle \\ &= \langle w, w\rangle \\ &\geq 0 \end{aligned}$$

because the scalar product is positive definite. Hence, $P$ is semi-positive.

Now, suppose $v, w \in V$, such that $v = w_1 + \widetilde{w_1}$ and $w = w_2 + \widetilde{w_2}$. We then have

$$\begin{aligned} &\langle w_1, w_2\rangle = \langle w_1, w_2\rangle \\ &\implies \langle w_1, w_2 + \widetilde{w_2}\rangle = \langle w_1 + \widetilde{w_1}, w_2\rangle \\ &\implies \langle Pv, w\rangle = \langle v, Pw\rangle \end{aligned}$$

and hence $P$ is symmetric.

**11.** Again, suppose $v, w \in V$ such that $v = w_1 + \widetilde{w_1}$ and $w = w_2 + \widetilde{w_2}$. We have

$$\begin{aligned} &\langle cw_1, w_2\rangle = \langle w_1, cw_2\rangle \\ &\implies \langle cw_1, w_2 + \widetilde{w_2}\rangle = \langle w_1 + \widetilde{w_1}, cw_2\rangle \\ &\implies \langle Av, w\rangle = \langle v, Aw\rangle \end{aligned}$$

and hence $A$ is symmetric.

**14.** Suppose $V$ is a finite dimensional vector space with a positive definite scalar product. Suppose $A$ is an operator such that

$$\langle Av, Aw\rangle = \langle v, w\rangle$$

for all $v, w \in V$. This means that ${}^tAA = I$, where $I$ is the identity operator. This means that the matrix of the operator ${}^tAA$ is the identity matrix, and hence

**Unitary Operators:** Let $V$ be a finite dimensional vector space over $\mathbb{R}$ with a positive definite scalar product. A map $A \in \mathrm{End}(V)$ is said to be *unitary* if for all $v, w$ in $V$, it is true that

$$\langle Av, Aw\rangle = \langle v, w\rangle.$$

Unitary maps are also called *orthogonal* or *norm-preserving* maps. Let's prove the following theorem which justifies the name 'unitary':

**Theorem 1.25.** Suppose $V$ is a finite dimensional vector space over $\mathbb{R}$ with a positive definite scalar product. Then, for $A \in \mathrm{End}(V)$, the following are equivalent:

(1) $A$ is unitary.
(2) $A$ is norm-preserving.
(3) $A$ maps unit vectors to unit vectors.

*Proof:* Let's first prove the equivalence between (1) and (2). First suppose $A$ is unitary. Then, for any $v \in V$, we have

$$\langle v, v\rangle = \langle Av, Av\rangle$$

and hence $A$ is norm-preserving. Conversely, suppose $A$ is norm preserving. Then, it follows that $\langle v, v\rangle = \langle Av, Av\rangle$ for all $v \in V$. Now, for $v, w \in V$, we have

$$\begin{aligned} 2\langle v, w\rangle &= \langle v+w, v+w\rangle - \langle v, v\rangle - \langle w, w\rangle \\ &= \langle Av+Aw, Av+Aw\rangle - \langle Av, Av\rangle - \langle Aw, Aw\rangle \\ &= 2\langle Av, Aw\rangle \end{aligned}$$

and hence we get that $\langle v, w\rangle = \langle Av, Aw\rangle$, which means that $A$ is unitary.

Now, let's prove the equivalence between (2) and (3). If (2) is true, then any unit vector $v$ will always be mapped to a unit vector $Av$, and hence (2) implies (3). Now, suppose (3) is true, and let $v$ be any vector in $V$. Then, observe that

$$\frac{v}{\sqrt{\langle v, v\rangle}}$$

is a unit vector. This means that

$$\frac{Av}{\sqrt{\langle v, v\rangle}}$$

is a unit vector. So, this means that

$$\left\langle \frac{Av}{\sqrt{\langle v, v\rangle}}, \frac{Av}{\sqrt{\langle v, v\rangle}} \right\rangle = 1$$

which means that

$$\langle Av, Av\rangle = \langle v, v\rangle$$

and hence $A$ is norm-preserving, and hence (3) implies (2) as well. This completes the proof.

There is also a connection between unitary maps and their transpose:

**Theorem 1.26.** Let $V$ be a finite dimensional vector space over $\mathbb{R}$ with a positive definite scalar product. $A \in \mathrm{End}(\mathrm{V})$ is unitary if and only if ${}^tAA = I$, where $I$ is the identity mapping.

*Proof:* Suppose $A$ is unitary. Then, for any vectors $v, w$ in $V$, we have

$$\langle Av, Aw \rangle = \langle v, w \rangle$$

which means that

$$\langle {}^t AAv, w \rangle = \langle v, w \rangle$$

and this implies that ${}^t AA = I$, where $I$ is the identity operator.

If we take $V = K^n$, and the scalar product to be the dot product, then the matrix $A$ is said to be *orthogonal* or *unitary* if ${}^t AA = I$, or

$${}^t A = A^{-1}$$

## Diagonalisation of Self Adjoint Maps

First, I would apologise if the content in this section is repeated in an earlier section (there is a slight chance, but most of the content is new). I wrote this section after almost three months of not touching the pdf.

Suppose $V$ is a finite dimensional vector space over $K$, where $K \in \{\mathbb{R}, \mathbb{C}\}$, and suppose $V$ is equipped with a positive-definite hermitian product (if $K = \mathbb{R}$ then the product will just be a scalar product).

Here is a general fact (general because the only condition on the inner product will be non-degeneracy): suppose the scalar (or hermitian) product is non-degenerate (positive-definite products are always non-degenerate). Then, the map $v \mapsto f_v$ is a linear **isomorphism** (and if $V$ is a space over $\mathbb{C}$, with a hermitian product, then it is an *anti-linear* isomorphism) between $V$ and $V^*$, where $f_v : V \to V$ is defined as

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

The above fact means that for non-degenerate products, every functional "looks like" the product taken with some fixed vector. Try proving this fact, it is not difficult. For such a space, let $A$ be an operator. Then, try to show that there is a unique operator $A^*$ on $V$ such that

$$\langle Av, w \rangle = \langle v, A^* w \rangle$$

for all $v, w \in V$. The operator $A^*$ is called the *adjoint* of $A$. $A$ is said to be *self-adjoint* if $A^* = A$. It is also easy to show that the adjoint satisfies the usual matrix conjugate transpose properties.

Now, let's come back to our original vector space $V$ with a positive definite product. In the following theorem, we will show how conjugate transpose is related to the adjoint:

**Theorem 1.27.** Suppose $\{e_1, ..., e_n\}$ is an orthonormal basis of $V$ (existence is clear because of positive-definiteness). Suppose $T$ is an operator, and let $A$ be the matrix of $T$ with respect to this basis. Then, the matrix of $T^*$ is $\overline{A}^t$, i.e the conjugate transpose.

*Proof:* Let $f_1, ..., f_n$ be the standard basis of $K^n$, and regard $A$ is an operator on $K^n$. Let $B$ be the matrix of $T^*$, and similarly we regard $B$ as an operator. Observe that $B_{ij}$ is given by

$$B_{ij} = \langle Bf_j, f_i \rangle$$

where the above inner-product is the standard one on $K^n$. Also, we know that

$$\langle Bf_j, f_i \rangle = \langle f_j, Af_i \rangle = \overline{A_{ji}}$$

which proves that $B = \overline{A}^t$

From the last theorem, we see that an operator is self-adjoint if and only if its matrix is hermitian (hermitian is the same as symmetric if $K = \mathbb{R}$).

We now define the *complexification* of an operator over $\mathbb{R}^n$. Suppose $V = \mathbb{R}^n$, and let the matrix of $T : V \to V$ be $A$. The complexification of $T$, denoted by $T_{\mathbb{C}}$, is the map $T_{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^n$ given by

$$T_{\mathbb{C}}(x) = Ax$$

where the last equation is matrix multiplication. In other words, we regard $T$ is an operator on $\mathbb{C}^n$. Now, it is easy to see that $T$ is self-adjoint if and only if $T_{\mathbb{C}}$ is self adjoint, because a real matrix is symmetric if and only if it is hermitian.

We are now in position to prove some very important results on self-adjoint maps:

**Theorem 1.28.** Suppose $V$ is a vector space over $K$ with a positive definite hermitian product, and let $T$ be a self-adjoint operator. Then, all the eigenvalues of $T$ are real. In particular, if $K = \mathbb{R}$, then $T$ has an eigenvector in $V$

*Proof:* First, suppose $K = \mathbb{C}$, so that $T$ is hermitian. Let $\lambda$ be an eigenvalue of $T$ (exists because $\mathbb{C}$ is algebraically closed). Let $v \neq O$ be a corresponding eigenvector. Then,

$$\lambda||v||^2 = \lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}||v||^2$$

and since $||v||^2 \neq 0$, we see that $\lambda$ is real.

Now, suppose $K = \mathbb{R}$, so that $T$ is symmetric. Let $A$ be the matrix of $T$ with respect to some basis of $V$. Consider the linear operator on $\mathbb{R}^n$, whose matrix with respect to the standard basis is $A$. Let $A_{\mathbb{C}}$ be the complexification of $A$. Since $A$ is real and symmetric, $A_{\mathbb{C}}$ hermitian, and hence it is self-adjoint. By the first paragraph of this proof, any eigenvalue of $A_{\mathbb{C}}$ is real. Since eigenvalues are precisely the roots of the characteristic polynomial, all roots of the characteristic polynomial are real. Finally, observe that the characteristic polynomials of $A_{\mathbb{C}}$ and $A$ are exactly the same (because they are the same matrices), and hence all eigenvalues of $A$ are real. This completes the proof.

So, it follows that any self-adjoint map has real eigenvalues. It actually turns out that any self-adjoint map is diagonalisable, which we will now prove. First, let us prove a lemma:

**Lemma:** Suppose $T$ is a self-adjoint operator on $V$. Let $\lambda$ be an eigenvalue of $T$, and consider the eigenspace $V_\lambda$. Then, $V_\lambda^\perp$ is $T$-invariant, and $T$ is self-adjoint on $V_\lambda^\perp$.

*Proof:* First, we will show that $V_\lambda^\perp$ is $T$-invariant. So, suppose $w \in V_\lambda^\perp$. Then, for any $v \in V_\lambda$, we have

$$\langle Tw, v\rangle = \langle w, Tv\rangle = \overline{\lambda}\langle w, v\rangle = 0$$

and hence $Tw \in V_\lambda^\perp$, so that $V_\lambda^\perp$ is $T$-invariant.

That $T$ is self-adjoint on $V_\lambda^\perp$ is clear because it is self-adjoint on $V$.

Finally, we can show that any self-adjoint operator is diagonalisable *with* an orthonormal basis:

**Theorem 1.29.** Let $T$ be self-adjoint. Then, there is an orthonormal basis of $V$ consisting of eigenvectors of $T$. Hence, $T$ is diagonalisable.

*Proof:* We proceed by induction on the dimension.

Let $\lambda$ be an eigenvalue of $T$ (even if the ground field is $\mathbb{R}$, an eigenvalue exists because of self-adjointness). Let $V_\lambda$ be the corresponding eigenspace. If $V = V_\lambda$, then we are done, because we can just choose an orthonormal basis of $V_\lambda$. If not, then

$$V = V_\lambda \oplus V_\lambda^\perp$$

and since $T$ restricted to $V_\lambda^\perp$ is self-adjoint, the claim follows by induction.

The last claim is a very powerful result, which is used in many areas of mathematics.

Next, we will discuss unitary maps, even though they are discussed in bit in earlier sections. The imporant takeaway will be that self-adjoint matrices (or maps) can be diagonalised via unitary matrices (or maps). We begin with the following theorem:

**Theorem 1.30.** Suppose $\Lambda : V \to V$ is a linear map. Then, the following are equivalent:

(1) $\Lambda$ is invertible and $\Lambda^{-1} = \Lambda^*$
(2) $\Lambda$ maps every orthonormal basis of $V$ to an orthonormal basis of $V$.
(3) $\Lambda$ transforms *some* orthonormal basis of $V$ to an orthonormal basis of $V$.
(4) $\Lambda$ preserves the inner product.

*Proof:* First, let us quickly see the equivalence between (1) and (4). Observe that if $A$ preserves the inner-product, then

$$\langle \Lambda v, \Lambda w\rangle = \langle v, w\rangle = \langle v, \Lambda^*\Lambda w\rangle$$

for all $v, w \in V$, and hence we see that $\Lambda^*\Lambda = I$, implying that $\Lambda^{-1} = \Lambda^*$. The converse is similarly proven.

(3) follows from (2) easily. To prove (2) from (3), suppose $\{v_1, ..., v_n\}$ (an orthonormal basis) is mapped to $\{v_1', ..., v_n'\}$ (another orthonormal basis), and let $\{w_1, ..., w_n\}$ be another orthonormal basis of $V$ (different from $\{v_1, ..., v_n\}$). Let $i, j \in \{1, ..., n\}$. Let

$$\begin{aligned} w_i &= a_{i1}v_1 + ... + a_{in}v_n \\ w_j &= a_{j1}v_1 + ... + a_{jn}v_n \end{aligned}$$

and hence we get that

$$\begin{aligned} \Lambda w_i &= a_{i1}v_1' + ... + a_{in}v_n' \\ \Lambda w_j &= a_{j1}v_1' + ... + a_{jn}v_n' \end{aligned}$$

and hence

$$\langle \Lambda w_i, \Lambda w_j\rangle = a_{i1}\overline{a_{j1}} + ... + a_{in}\overline{a_{jn}} = \langle w_i, w_j\rangle = \delta_{ij}$$

and hence $\{\Lambda w_1, ..., \Lambda w_n\}$ is an orthonormal basis of $V$. This proves both directions.

Finally, we will look at the equivalence between (2) and (4). If (4) is true and that $\Lambda$ preserves inner product, then (2) clearly follows. Next, suppose (2) is true. Let $\{v_1, ..., v_n\}$ be an orthonormal basis of $V$, and hence $\{\Lambda v_1, ..., \Lambda v_n\}$ is an orthonormal basis of $V$. Let $v, w \in V$, so that

$$\begin{aligned} v &= a_1v_1 + ... + a_nv_n \\ w &= b_1v_1 + ... + b_nv_n \end{aligned}$$

Now, we have

$$\begin{aligned} \langle \Lambda v, \Lambda w\rangle &= \langle a_1\Lambda v_1 + ... + a_n\Lambda v_n, b_1\Lambda v_1 + ... + b_n\Lambda v_n\rangle \\ &= a_1\overline{b_1} + ... + a_n\overline{b_n} \\ &= \langle v, w\rangle \end{aligned}$$

and hence $\Lambda$ preserves inner products. This completes the proof.

A matrix $A$ is said to be unitary if $A^{-1} = \overline{A^t}$.

**Some properties of unitary maps:** Suppose $\Lambda$ is a unitary operator on $V$, and let $A$ be its matrix.

(1) Clearly, $A$ satisfies $\det A^2 = 1$, and hence $\det A$ lies on the unit circle.
(2) Suppose $\lambda$ is an eigenvalue of $\Lambda$. Let $v$ be a corresponding unit eigenvector. Then we have

$$1 = \langle \Lambda v, \Lambda v \rangle = |\lambda|^2 ||v||^2 = 1$$

and hence $|\Lambda|^2 = 1$. So again, $\Lambda$ lies on the unit circle.

We will now do a spectral theorem for unitary maps as well.

## 2. Positive Definite Operators

As before , suppose $V$ is a finite dimensional vector space over $K$.

A self-adjoint operator $T$ on $V$ is said to be *positive semi-definite* if $\langle Tv, v \rangle \geq 0$ for all $v \in V$. $T$ is said to be *positive definite* if the inequality is strict for all $v \neq O$. Similarly *negative semi-definite* and *negative definite* operators are defined.

These operators are characterised by the sign of the eigenvalues:

**Theorem 2.1.** Let $T$ be a self-adjoint operator. Then, $T$ is positive semi-definite (or positive definite) if and only if all eigenvalues of $T$ are non-negative (or strictly positive).

*Proof:* First, suppose $T$ is positive semi-definite (or positive definite). Let $\lambda$ be an eigenvalue. Let $v$ be a corresponding eigenvector. Then, we have

$$0 \leq \langle Tv, v \rangle = \lambda \langle v, v \rangle$$

and hence $\lambda \geq 0$ (inequality is strict if $T$ is positive definite). Conversely, suppose all eigenvalues are non-negative (or positive). Let $\{v_1, .., .v_n\}$ be a spectral basis (exists because $T$ is self-adjoint). Let $v \in V$, and let

$$v = a_1 v_1 + ... + a_n v_n$$

Then,

$$\begin{aligned}
\langle Tv, v \rangle &= \langle a_1 \lambda_1 v_1 + ... + a_n \lambda_n v_n, a_1 v_1 + ... + a_n v_n \rangle \\
&= |a_1|^2 \lambda_1 + ... + |a_n|^2 \lambda_n \\
&\geq 0
\end{aligned}$$

(the inequality is strict if $T$ is positive definite). This completes the proof.

**Note:** The analogous fact is also true for negative semi-definite operators.

The following theorem is a special characterisation for dimension 2:

**Theorem 2.2.** Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ be a symmetric $2 \times 2$ matrix. Then, $A$ is positive definite if and only if $\det A > 0$ and $a > 0$. It is negative definite if and only if $\det A > 0$ and $b < 0$.

*Proof:*

## 3. Bilinear Maps and Quadratic Forms

Suppose $U, V$ and $W$ are vector spaces. A map $g : U \times V \to W$ is said to be *bilinear* $g(x, y)$ is linear in both arguments (similarly, multilinear maps are defined). Scalar products are examples of bilinear maps. An example of multilinear map is the determinant.

In this section, we will be interested in bilinear *forms* (i.e, maps taking values in the ground field).

First, suppose we are given an $m \times n$ matrix $A$ over $K$. Define a map $g_A : K^m \times K^n \to K$ given by

$$g_A(X, Y) = X^t AY$$

It is easy to see that $g_A$ is then a bilinear form. We now prove the following theorem:

**Theorem 3.1.** Given a bilinear form $g : K^m \times K^n \to K$, there is a *unique* matrix $A$ such that $g = g_A$, i.e

$$g(X, Y) = X^t AY$$

The set of bilinear maps from $K^m \times K^n$ to $K$ is a vector space over $K$, denoted by $\text{Bil}(K^m \times K^n, K)$ and the association

$$A \to g_A$$

gives an isomorphism between $M_{m\times n}(K)$ and $\text{Bil}(K^m \times K^n, K)$.

*Proof:* First, suppose a bilinear form $g$ is given to us. Fix the standard basis $\{E_1, ..., E_m\}$ and $\{E'_1, ..., E'_n\}$ of $K^m$ and $K^n$.

Now, let $X = c_1E_1 + ... + c_mE_m$, and $Y = d_1E'_1 + ... + d_nE'_n$. Then, we have

$$\begin{aligned} g(X, Y) &= g(c_1E_1 + ... + c_mE_m, d_1E'_1 + ... + d_nE'_n) \\ &= \sum_{i=1}^{m} c_i g(E_i, d_1E'_1 + ... + d_nE'_n) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j g(E_i, E_j) \end{aligned}$$

Now, let $A_{ij} = g(E_i, E_j)$, so that $A$ is an $m \times n$ matrix. The above equation says that

$$g(X, Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j A_{ij} = X^t AY$$

and hence the required matrix has been found. Showing the uniqueness of $A$ is not difficult, and I will skip that. Observe that we could have fixed some other basis of $K^n$ and $K^m$ as well.

So, the map $A \to g_A$ from $M_{m\times n}(K)$ to $\text{Bil}(K^m \times K^n, K)$ is clearly one-one and onto. To show that it is a homomorphism, observe that if $g_1$ and $g_2$ are two bilinear forms, then we have

$$(g_1 + g_2)(X, Y) = g_1(X, Y) + g_2(X, Y) = X^t A_1 Y + X^t A_2 Y = X^t (A_1 + A_2)Y$$

and hence the matrix of $g_1 + g_2$ is $A_1 + A_2$. This shows the isomorphism.

The two exercises in the next section are important properties connecting inner-products to bilinear forms.

## 4. V,§4. Exercises

This is the solution to problem **1.** and **2.** combined.

First, suppose $A$ is an $n \times n$ symmetric matrix. Let $g_A$ be the associated bi-linear form. We have that $g_A(X, Y) = X^t AY$ and $g_A(Y, X) = Y^t AX$. Now, observe that $X^t AY$ is the standard dot product of $X$ and $AY$, which we write as $X \cdot AY$ (note that this dot product is over any arbitrary field). This is commutative, and hence

$$X^t AY = X \cdot AY = AY \cdot X = (AY)^t X = Y^t AX$$

and hence $g_A(X, Y) = g_A(Y, X)$.

Conversely, suppose $g_A(X, Y) = g_A(Y, X)$ for all $X, Y \in K^n$. Let $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$. We expand both sides, and get

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} X_i Y_j = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} X_j Y_i$$

Now, take for any $1 \leq a, b \leq n$, let $X$ be the vector whose $a^{\text{th}}$ coordinate is 1, and rest are 0. Similarly, let $Y$ be the vector whose $b^{\text{th}}$ coordinate is 1 and rest are 0. So, we have

$$A_{ab} = A_{ba}$$

and hence $A$ is symmetric.

It is now easy to see that for a symmetric matrix $A$, the bilinear form associated with it defines an inner product.

**Positive definite Inner Products.** : Let $V$ be a finite dimensional vector space with a positive definite inner (scalar or hermitian) product over $K$, where $K \in \{\mathbb{R}, \mathbb{C}\}$.

By a similar strategy as we did in **Theorem 3.1**, we will see a one-to-one correspondance between positive definite operators and positive definite inner products on $V$.

First, suppose $T : V \to V$ is a positive definite operator. Define $\langle v, w \rangle_T = \langle Tv, w \rangle$. Then, it is not hard to see that $\langle \cdot \rangle_T$ is an inner product on $V$. Conversely, suppose an inner product $\langle \cdot \rangle'$ is given on $V$. Let $\{v_1, ..., v_n\}$ be an orthogonal basis of $V$. Observe that if $a_1 v_1 + ... + a_n v_n$ and $b_1 v_1 + ... + b_n v_n$ are any two vectors, then we have

$$\langle a_1 v_1 + ... + a_n v_n, b_1 v_1 + ... + b_n v_n \rangle' = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{b_j} \langle v_i, v_j \rangle'$$

So, we define a linear map $T$ such that the entries of the matrix of $T$ (with respect to the given basis) are $\langle v_j, v_i \rangle'$. In other words, $T$ is defined by

$$T(v_j) = \sum_{i=1}^{n} \langle v_j, v_i \rangle' v_i$$

Then, observe that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{b_j} \langle v_i, v_j \rangle' = \langle Tv, w \rangle$$

and hence it follows that

$$\langle v, w \rangle' = \langle Tv, w \rangle$$

and hence $T$ is a positive definite operator.

**Quadratic Forms.** A map $Q : V \to \mathbb{R}$ is said to be a quadratic form if there exists a self-adjoint operator $T : V \to V$ such that

$$Q(v) = \langle Tv, v \rangle$$

Note that we can define quadratic forms over other fields as well, but we restrict ourselves to $\mathbb{R}$ for now. Take an orthogonal basis $\{f_1, ..., f_n\}$ of $V$, and suppose $A$ is the matrix of $A$ with respect to this basis (so that $A$ is symmetric). Then, we see that

$$\begin{aligned} Q(x_1 f_1 + ... + x_n f_n) &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \langle T v_i, v_j \rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j A_{ij} \\ &= \sum_{i=1}^{n} A_{ii} x_i^2 + \sum_{i<j} 2 A_{ij} x_i x_j \end{aligned}$$