Year 2 — Linear Algebra

Based on lectures by Dr Mark Callaway

Notes taken by James Arthur

Summer Term 2021

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Contents

1 Introduction to Linear Algebra

In this definition we will define some important things informally and then later come back with formal definitions. Firstly we talk about how we know that a function or relationship is linear,

Definition 1.1. *(Linearity)* We define for $f : X \to Y$ to say it's linear iff,

$$f(x_1 + x_2) = f(x_1) + f(x_2)$$
$$f(ax) = af(x)$$

Now we are going define what a vector space is very loosely,

Definition 1.2. *(Vector Space)* Let V be a set and let it be defined by two operations, addition and scalar multiplication.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$
$$a\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix}$$

We can go and take this loose definition and now define what subspaces are,

Definition 1.3. *(Subspace)* A subspace of a vectorspace V is a subset $W \subset V$ that still has a vector structure.

We now take the final ideas and loose definitions we need to get a gripse on the topic,

Definition 1.4. *(Dimension)* The number of independent directions or 'degrees of freedom'.

Definition 1.5. *(Linear Transformation)* Let V and W be vectorspaces over the same field, $\mathbb{F}$. A linear transformation $T : V \to W$ preserves structure of V,

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) \qquad \forall\, \mathbf{v}_1, \mathbf{v}_2 \in V$$
$$T(a\mathbf{v}) = aT(\mathbf{v}) \qquad \forall a \in \mathbb{F}, \mathbf{v} \in V$$

2 Vector Spaces

Before we defined Vector spaces informally, now let us define it formally.

Definition 2.1. Let V be a set, $\mathbb{F}$ be a field of scalars, $\oplus$ be a binary operation on V, that is a function $\oplus : V \times V \to V$, which we call addition and let $\odot : \mathbb{F} \times V \to V$ be another function which we call scalar multiplication. Then we have eight axioms on $(V, \oplus, \odot)$, [1]

$$\forall\, \mathbf{u}, \mathbf{v} \in V, \mathbf{u} \oplus \mathbf{v} = \mathbf{v} \oplus \mathbf{u} \qquad \texttt{add\_comm}$$
$$\forall\, \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w}) = (\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w} \qquad \texttt{add\_assoc}$$
$$\exists\, \mathbf{0}_v, \forall\, \mathbf{u} \in V, \mathbf{u} \oplus \mathbf{0}_v = \mathbf{u} = \mathbf{0}_v \oplus \mathbf{u} \qquad \texttt{add\_id}$$
$$\forall\, \mathbf{u} \in V, \exists\, \mathbf{x} \in V, \mathbf{u} \oplus \mathbf{x} = \mathbf{x} \oplus \mathbf{u} = \mathbf{0}_v \qquad \texttt{add\_inv}$$
$$\forall\, a \in \mathbb{F}, \forall\, \mathbf{u}, \mathbf{v} \in V, a \odot (\mathbf{u} \oplus \mathbf{v}) = (a \odot \mathbf{u}) \oplus (a \odot \mathbf{v}) \qquad \texttt{left\_dist\_sc\_add}$$
$$\forall\, a, b \in \mathbb{F}, \forall\, \mathbf{u} \in V, (a + b) \odot \mathbf{u} = (a \odot \mathbf{u}) \oplus (b \odot \mathbf{u}) \qquad \texttt{right\_dist\_sc\_add}$$
$$\forall\, a, b \in \mathbb{F}, \forall\, \mathbf{u} \in V, a \odot (b \odot \mathbf{u}) = ab \odot \mathbf{u} \qquad \texttt{sc\_mul\_mul\_sc}$$
$$\forall\, \mathbf{u} \in V, 1 \odot \mathbf{u} = \mathbf{u} \qquad \texttt{mul\_id}$$

```
class vector_space' (R : Type u) (V : Type v) [field R] [add_comm_group V] extends
    has_scalar R V :=
  (add_comm : ∀ (u v : V), u + v = v + u)
  (add_assoc : ∀(u v w : V), (u + v) + w = u + (v + w))
  (add_id_right : ∀ (u : V), u + (0 : V) = u)
  (add_id_left : ∀ (u : V), (0 : V) + u = u)
  (add_inv_left : ∀ (u : V), ∃ (x : V), x + u = (0: V))
  (add_inv_right : ∀ (u : V), ∃ (x : V), u + x = (0: V))
  (left_dist : ∀ (a : R) (u v :  V), a · ( u + v ) = a · u + a · v)
  (right_dist : ∀ (a b : R) (u : V), (a + b) · u = a · u + b · u)
  (smul_smul : ∀ (a b : R) (u : V), a · (b · u) = (a * b) · u)
  (smul_id : ∀ (u : V), (1 : R) · u = u)
```

We have these vector spaces with respect to addition, this is because addition is a binary operation and that something like the inner product isn't. The inner product is a bilinear form.

Definition 2.2. *(Bilinear Form)* Let V be a vector space, and define $\otimes : V \times V \to \mathbb{F}$, where $\mathbb{F}$ is a field of scalars, such that,

$$\forall\, \mathbf{u}, \mathbf{v}, \mathbf{w}, \in V, (\mathbf{u} \oplus \mathbf{v}) \otimes \mathbf{w} = (\mathbf{u} \otimes \mathbf{w}) \oplus (\mathbf{u} \otimes \mathbf{v})$$
$$\forall\, \mathbf{u}, \mathbf{v}, \mathbf{w}, \in V, \mathbf{u} \otimes (\mathbf{v} \oplus \mathbf{w}) = (\mathbf{u} \otimes \mathbf{v}) \oplus (\mathbf{u} \otimes \mathbf{w})$$
$$\forall a \in \mathbb{F}, \forall \mathbf{u}, \mathbf{v} \in V, (a\mathbf{u}) \otimes \mathbf{v} = \mathbf{u} \otimes (a\mathbf{v}) = a(\mathbf{u} \otimes \mathbf{v})$$

and we define Euclidian space so we can use 'normal' notation.

Definition 2.3. *(Euclidian Space)* The vector space $(\mathbb{R}^n, \oplus, \odot)$ with $\oplus$ and $\odot$ defined in the familiar way, is defined as the Euclidian space $\mathbb{R}^n$.

Remark. A vectorspace over the field of reals is called a real vector space and over the complexes a complex vector space.

Definition 2.4. *(Closure)* A set S is closed under a binary operation if applying the operation to arbitrary elements of S yields an element of S

[1] We should technically have a quadruple $(V, \mathbb{F}, \oplus, \odot)$. Where we don't just assume the vector field is just over a certain field.

Hence we need a vectorspace to be,

(i) Closed under addition ($\oplus$)

(ii) Closed under scalar multiplication ($\odot$)

Checking closure effectively means that we are checking the codomain of $\oplus$ and $\odot$ are correct.

Lemma 2.5. The additive identity of a vector space is unique

Proof. Suppose the additive identity isn't unique, so $\exists \mathbf{z} \in V, \mathbf{z} \oplus \mathbf{v} = \mathbf{v}\ \forall \mathbf{v} \in V$. Since $\mathbf{0}_v$ is an additive identity we have,

$$\mathbf{0}_v \oplus \mathbf{z} = \mathbf{z}$$
$$\mathbf{0}_v \oplus \mathbf{z} = \mathbf{0}_v$$

Then,

$$\mathbf{z} = \mathbf{0}_v \oplus \mathbf{z} = \mathbf{0}_v$$

and so the additive identity is unique. □

Lemma 2.6. Each element of a vector space has a unique additive inverse

Proof. Let $\mathbf{u} \in V$. Suppose both $\mathbf{v}, \mathbf{w} \in V$ are the additive inverse of $\mathbf{u}$, hence,

$$\mathbf{u} \oplus \mathbf{v} = \mathbf{0}_v \qquad \mathbf{u} \oplus \mathbf{w} = \mathbf{0}_v$$

Then,

$$\begin{aligned} \mathbf{v} &= \mathbf{0}_v \oplus \mathbf{v} \\ &= (\mathbf{w} \oplus \mathbf{u}) \oplus \mathbf{v} \\ &= \mathbf{w} \oplus (\mathbf{u} \oplus \mathbf{v}) \\ &= \mathbf{w} \oplus \mathbf{0}_v \\ &= \mathbf{w} \end{aligned}$$

□

Lemma 2.7. Let $(V, \oplus, \odot)$ be a vector space. Then $0 \odot \mathbf{u} = \mathbf{0}_v \quad \mathbf{u} \in V$.

Proof.

```
lemma zero_smul' (u : V) : (0 : V) = (0 : R) · u :=
begin
  have H1 : (0 : R) · u = ((0 : R) · u) + ((0 : R) · u),
  { conv_lhs {
    rw ← zero_add (0 : R),
    rw right_dist},
  },
  have H2 : ((0 : R) · u) + -((0 : R) · u) = (0 : V) := by rw add_neg_self,
  conv_lhs at H2 {
    congr,
    rw H1,
  },
  rw [add_assoc, add_neg_self, add_zero] at H2,
  exact H2.symm,
end
```

□

Lemma 2.8. Let $(V, \oplus, \odot)$ be a vector space. Then $a \odot \mathbf{0}_v = \mathbf{0}_v \quad \forall\, a \in \mathbb{F}$.

Proof.
```
lemma smul_zero' (u : V) : a · (0 : V) = (0 : V) :=
begin
  have H1 : a · (0 : V) = (a · (0 : V)) + (a · (0 : V)),
  { conv_lhs { rw ← zero_add (0 : V),
    rw left_dist},
  },
  have H2 : (a · (0 : V)) + -(a · (0 : V)) = (0 : V) := by rw add_neg_self,
  conv_lhs at H2{
    congr,
    rw H1,
  },
  rw [add_assoc, add_neg_self, add_zero] at H2,
  exact H2,
end
```

□

Lemma 2.9. Let $(V, \oplus, \odot)$ be a vector space. Then $(-1) \odot \mathbf{u} = -\mathbf{u} \quad \forall\, \mathbf{u} \in V$.

Proof.
```
lemma inv_eq_neg (u : V) (x : R) : -u = (-1 : R) · u :=
begin
  rw ← add_left_inj u,
  rw neg_add_self,
  have H : (-1 : R) · u + ((1 : R) · u) = (-1 : R) · u + u := by rw smul_id,
  rw ← H,
  rw ← right_dist,
  rw neg_add_self,
  rw zero_smul' u,
end
```

□

Definition 2.10. *(Subspace)* Let $(V, \oplus, \odot)$ be a vector space and let $U \subset V$. We say $(U, \oplus, \odot)$ is a subspace of $(V, \oplus, \odot)$ if $(U, \oplus, \odot)$ is a vector space.

Lemma 2.11. Let $(V, \oplus, \odot)$ be a vector space and $U \subset V$. The triple $(U, \oplus, \odot)$ is a subspace of $(V, \oplus, \odot)$ $\iff$ U has,

(i) $\mathbf{0}_v \in U$

(ii) $\mathbf{u} \oplus \mathbf{w} \in U \qquad \forall\, \mathbf{u}, \mathbf{w} \in U$

(iii) $a \odot \mathbf{u} \in U \qquad \forall\, a \in \mathbb{F}, \mathbf{u} \in U$

Lemma 2.12. Let V be a vector space and U and W be subspaces of V, then $U \cap W$ is a subspace of V.

Proof. To prove this lemma all we need to do is check each of the subspace axioms for our subspace $U \cap W$, given we know that U and W are subspaces themselves. Hence we first turn our attention to the zero vector,

- As we know that U and W are subspaces of V, they must both contain $\mathbf{0}_v$ (the zero vector of V). Hence the intersection of the subspaces must also contain the zero vector. $\mathbf{0}_v \in U \cap W$.
- If we now take a $\mathbf{u}$ and $\mathbf{v}$ that are in both U and W, then we can say that $\mathbf{u}+\mathbf{v} \in U$ and $\mathbf{u}+\mathbf{v} \in W$, as they are both closed under addition. Hence, $\mathbf{u}+\mathbf{v} \in U \cap W$, and so $U \cap W$ is also closed under addition.
- Finally we take a $\mathbf{w}$ that is in both U and W, now we say that if $a \in \mathbb{F}$ and is an arbitrary scalar, that $a\mathbf{w} \in U$ and also $a\mathbf{w} \in W$ as they are both closed under smul. Hence, $a\mathbf{w} \in U \cap W$ and $U \cap W$ is closed under smul.

As we have all of these properties holding $U \cap W$ is a subspace of V. □

Lemma 2.13. Let V be a vector space and U and W be subspaces of V, then $U \cup W$ is a subspace of V if and only if $U \subset W$ or $W \subset U$.

Proof. Firstly we prove the forward direction, for this we go for a proof by contradiction, so we assume that $W \not\subset U$ (which is symmetrical to the other case). Now we let $\mathbf{x} \in W$ and $\mathbf{y} \in W \setminus U$. Hence $\mathbf{x} \in U \cup W$ and $\mathbf{y} \in U \cup W$. Hence by definition of set union, $\mathbf{x}+\mathbf{y} \in U$ and $\mathbf{x}+\mathbf{y} \in W$.

If $\mathbf{x}+\mathbf{y} \in W$, then $\mathbf{x}+\mathbf{y}+(-1)\mathbf{x} \in W$ and hence $\mathbf{y} \in W$. However, $\mathbf{y} \in W \subset U$, so $\mathbf{x}+\mathbf{y} \in U$. Then we can say $x \in U$. Hence $U \subset W$.

For the backward direction, notice that if $U \subset W$, then $U \cup W = W$, and symmetrically, hence $U \cup W$ is a subset of V. □

Definition 2.14. *(Sum of subsets)* Let V be a vector space, $U, W \subset V$, we define $U+W$ to be,

$$U + W = \{\mathbf{u}+\mathbf{w} : \mathbf{u} \in U \text{ and } \mathbf{w} \in W\}$$

Lemma 2.15. If U and W are subspaces of the vectorspace V, then the sum $U+W$ is a subspace of V

Proof. `add_id`: We know $\mathbf{0}_v \in U, W$, then $\mathbf{0}_v = \mathbf{0}_v + \mathbf{0}_v \in U+W$. Hence we have an additive identity in $U+W$.
`add_closure`: Take an arbitrary $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$. Now let $\mathbf{v}_1, \mathbf{v}_2 \in V$. So,

$$\mathbf{u}_1 + \mathbf{w}_1 = \mathbf{v}_1 \qquad \mathbf{u}_2 + \mathbf{w}_2 = \mathbf{v}_2$$

Now consider,

$$\begin{aligned}\mathbf{v}_1 + \mathbf{v}_2 &= (\mathbf{u}_1 + \mathbf{w}_1) + (\mathbf{u}_2 + \mathbf{w}_2) \\ &= (\mathbf{u}_1 + \mathbf{u}_2) + (\mathbf{w}_1 + \mathbf{w}_2) \in U + W\end{aligned}$$

Hence, $U+W$ is closed under addition.
`smul_closure`: Let $\mathbf{v} \in U+W$, then $\mathbf{v} = \mathbf{u}+\mathbf{w}$ for some $\mathbf{u} \in U$ and $\mathbf{w} \in W$. Then consider for some $a \in \mathbb{F}$,

$$\begin{aligned}a\mathbf{v} &= a(\mathbf{u}+\mathbf{w}) \\ &= a\mathbf{u} + a\mathbf{w} \in U + W\end{aligned}$$

Hence $U+W$ is closed under smul and $U+W$ is a vector space □

3 Linear (In)dependence, Spans and Basis

3.1 Linear Independence

Definition 3.1. (*List*) Let X be a set. A list of elements in X is a finite sequence if n elements of X, written $[x_1, x_2, \dots, x_n]$ where $x_1, \dots, x_n \in X$.

Note: We permit [ ] for $X = \varnothing$ as this makes sense.

Definition 3.2. (*Linear Combination*) Let V be a vectorspace and let $S = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ be a list of elements of V. A linear combination of elements of the list S is a vector of the form,

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \qquad a_1, \dots, a_n \in \mathbb{R}$$

If $S = [\quad]$, then we define the linear combination of S to be $\mathbf{0}_v$.

Definition 3.3. (*Linear Combination of a set*) Let V be a vectorspace and let S be subset of V. A linear combination of elements of the set S is,

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \qquad a_1, \dots, a_n \in \mathbb{R}$$

where $[\mathbf{v}_1, \dots, \mathbf{v}_n]$ is a list of elements of S.

Again, if $S = \varnothing$ the only linear combination is $\mathbf{0}_v$.

3.2 Span

Definition 3.4. (*Span*) Let V be a vectorspace and let S be a subset or list of vectors of V. The set of all linear combinations of S is called the span of S, $\operatorname{span} S$

If $V = \operatorname{span} S$, we say that S is the spanning set or list for V.

Lemma 3.5. Let V be a vectorspace and $S \subset V$. Then $\operatorname{span} S$ is always a subspace of V.

Proof. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a subset of vectors from V, then we know that $\operatorname{span} S$ is just all of the linear combinations of the elements of S, so,

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \in \operatorname{span} S$$

where $a_i \in \mathbb{R}$. We need to show that this is a subspace of V, so we need to show it satisfies the three subspace axioms,

– Showing that $\mathbf{0}_v \in \operatorname{span} S$, is rather simple, just let all the a_i's be 0. Then,

$$0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n = \mathbf{0}_v \in \operatorname{span} S$$

– Then we take arbitrary $\mathbf{u}_1, \mathbf{u}_2 \in \operatorname{span} S$ and we prove closure of addition,

$$\begin{aligned}\mathbf{u} + \mathbf{v} &= (a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) + (b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n) \\ &= (a_1 + b_1)\mathbf{v}_1 + \cdots + (a_n + b_n)\mathbf{v}_n \\ &\in \operatorname{span} S\end{aligned}$$

– Finally we show closure under scalar multiplication. Let $a \in \mathbb{F}$ and $\mathbf{u} \in \operatorname{span} S$. Then we can say,

$$\begin{aligned}a\mathbf{u} &= a(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= ac_1\mathbf{v}_1 + \cdots + cv_n\mathbf{v}_n \\ &\in \operatorname{span} S\end{aligned}$$

Hence as we have satisfied all of the axioms, $\operatorname{span} S$ is a subset of V if $S \subset V$. $\square$

Lemma 3.6. Let $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of vectors of $\mathbb{F}^m$. Let A be the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then the following are equivilent,

(i) $\operatorname{span} S = V$

(ii) The equation $A\mathbf{x} = \mathbf{v}$ has a solution $\forall\, \mathbf{v} \in V$.

(iii) Any echelon form of A has a pivot in every row.

Definition 3.7. (*Linearly (in)dependent*) Let V be a vectorspace and let $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of elements of V. If $\exists\, a_1, \ldots, a_n \in \mathbb{R}$, $a_1, \ldots, a_n \neq 0$ and, $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}_v$ then S is linearly dependent.

If S is not linearly dependent, S is linearly independent.

Definition 3.8. (*Linear Independence on sets*) We say S is linearly independent if every list of distinct elements of S is linearly independent. Otherwise S is linearly dependent.

Theorem 3.9. Let V be a vectorspace and $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of vectors from V with $n > 1$. Then S is linearly dependent iff one of the $\mathbf{v}_i$ can be written as the linear combination of the other.

Lemma 3.10. Let $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of vectors from $\mathbb{F}^m$. Let A be the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then the following are equivilent,

(i) S is linearly independent

(ii) The equation, $A\mathbf{x} = \mathbf{0}_{\mathbb{F}^m}$

(iii) Any echelon form of A has a pivot in every column.

Remark. $\operatorname{span} S = \mathbb{F}^n$ iff echelon form has a pivot in every row
S is linearly independent iff echelon form has a pivot in every column.

3.3 Bases

Definition 3.11. (*Basis*) Let V be a vectorspace and let S be a list/set of vectors from V. Then S is said to be a basis if it's spanning and linearly independent.

Definition 3.12. (*Standard Basis*) The ordered basis $[\mathbf{e}_1, \ldots, \mathbf{e}_n]$ where $\mathbf{e}_i$ is the vector with all 0's except a 1 in the ith position.

Remark. A basis has the same dimension as the vectorspace.

Lemma 3.13. Let $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of vectors from $\mathbb{F}^m$. Let A be the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then the following are equivilent,

(i) S is a spanning set for $\mathbb{F}^m$

(ii) S is linearly independent

(iii) S forms a basis for $\mathbb{F}^n$

(iv) $A\mathbf{x} = \mathbf{v}$ has a unique solution $\forall\, \mathbf{v} \in V$

(v) Any A_{EF} has a pivot in every row and column

(vi) $A_{RRE} = I$

(vii) A is invertible

(viii) $\det A \neq 0$

Theorem 3.14. $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of vectors of $\mathbb{F}^m$

(i) If $n < m$, then S doesn't span $\mathbb{F}^m$.

(ii) If $n > m$, then S is linearly dependent.

(iii) If S is a basis for $\mathbb{F}^m$, then $n = m$.

Proof. We know that S spans a vector space if $\operatorname{span} S = V$. Here we are considering S itself instead of $\operatorname{span} S$, hence S spans if $S = V$. By definition of set equality, each element in S must be in V, hence it follows for $S = V$, $n = m$ and if $n < m$, then S can't span V.

Then a vector space, V, only has k linearly independent basis (or spanning) vectors, where $k = \dim V$. Here $k = m$ hence, if there are n vectors ($n > m$) that supposedly span the space they are not all linearly independent, hence they are linearly dependent. □

Theorem 3.15. Let V be a vectorspace and let $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be a list of vectors of V. The list S is an ordered basis of V iff $\forall\, \mathbf{v} \in V, \exists\, a_i, i \in \mathbb{N}$,

$$\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{v}_i$$

Proof. First suppose that S is a basis of V. Then $\operatorname{span} S = V$. Thus any $\mathbf{v} \in V$ can be expressed in $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. Suppose that also, $\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$. Then,

$$\mathbf{v} - \mathbf{v} = (a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n = \mathbf{0}_v$$

This then implies that $a_i - b_i = 0 \quad \forall\, i \in \mathbb{N}$, hence $\mathbf{v}$ is unique.
Conversely suppose that $\forall\, \mathbf{v} \in V$ has a unique expression of the form,

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

by definition then $V = \operatorname{span} S$. Now suppose, $\mathbf{0}_v = a_1\mathbf{v}_1 + \ldots a_n\mathbf{v}_n$, by this we also have that $\mathbf{0}_v = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n$. By uniqueness, we must have $a_1 = \cdots = a_n = 0$, hence linearly independent. □

Definition 3.16. (*Coordinate Vector*) Let V be a vectorspace over a field $\mathbb{F}$ with ordered basis $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, then by thm 5.8, $\forall \mathbf{v} \in V, \mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ for unique scalars, $a_i \in \mathbb{F}, \quad i \in \mathbb{N}$. The coordinates vector of $\mathbf{v}$ wrt basis S is,

$$[\mathbf{v}]_S = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Theorem 3.17. Let V be a vectorspace with ordered basis S.

(i) If $\mathbf{u}, \mathbf{v} \in V$ and $a \in \mathbb{F}$, then $[\mathbf{u} + \mathbf{v}]_S = [\mathbf{u}]_S + [\mathbf{v}]_S$ and $[a\mathbf{u}]_S = a[\mathbf{u}]_S$

(ii) Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $\mathbf{w}$ be elements of V. Then $\mathbf{w}$ is a linear combination of $\mathbf{v}_i$ iff $[\mathbf{w}]_S$ is a linear combination of $[\mathbf{v}_i]_S$'s.

3.4 Isomorphism

Definition 3.18. (*Isomorphism*) Let V and W be vectorspaces over the same field $\mathbb{F}$. An isomorphism $V \to W$ is a bijective map, $T : V \to W$ st, $\forall\, \mathbf{u}, \mathbf{v} \in V$ and $a \in \mathbb{F}$ we have,

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \qquad T(a\mathbf{u}) = aT(\mathbf{u})$$

if $\exists$ an isomorphism $V \to W$, then V and W are said to be isomorphic, $V \cong W$

Corollary 3.19. If V is a vectorspace over $\mathbb{F}$ with an ordered basis $S = [v_1, \ldots, v_n]$ then the coordinate vector map $v \mapsto [\mathbf{v}]_S$ is an isomorphism from $V \to \mathbb{F}^n$. Hence, $V \cong \mathbb{F}^n$.

4 Dimensions and Direct Sums

Note from here on in, V is a vectorspace and I shall note whether it has infinite dimension. If I use U and W, these are subspaces of V.

4.1 Dimension

Corollary 4.1. Let $S = [\mathbf{v}_1, \dots, \mathbf{v}_m]$ be a basis for V and suppose $R = [\mathbf{u}_1, \dots, \mathbf{u}_n]$ is another list of vectors in V, then,

(i) If $n < m$ then R doesn't span V

(ii) If $n > m$, then R is linearly independent

(iii) If R is a basis for V, then $n = m$

(iv) If $n = m$ and R is linearly independent, then R is a basis for V.

(v) If $n = m$ and R spans V, then R is a basis for V.

Definition 4.2. (*Dimension*) Dimension is defined to be the size of any basis of V

Remark. Given Cor. 4.1 we can say that every dimension is uniquely defined as basis are unique.

4.2 Direct Sums

Theorem 4.3. $\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$

Definition. (*Direct Sum*) If $U + W = V$ and $U \cap W = \{\mathbf{0}_v\}$, then V is said to be the direct sum of U and W, denoted, $V = U \oplus W$.

Corollary 4.4. If $V = U \oplus W$, then,

$$\dim V = \dim U + \dim W$$

Theorem 4.5. $V = U \oplus W$ if and only if, $\forall\, \mathbf{v} \in V$ can be decomposed as,

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \qquad \forall\, \mathbf{u} \in U, \forall\, \mathbf{w} \in W$$

If $V = U \oplus W$, then, by thm 4.5, we can say $\forall\, \mathbf{v} \in V$ we can uniquely decompose it into $\mathbf{u} + \mathbf{w}$, $\forall\, \mathbf{u} \in U$ and $\forall\, \mathbf{w} \in W$. This leads us to projections,

Definition 4.6. (*Projection Mapping*) The projection onto U along W is the mapping, $P_u : V \to V$ defined by $P_u\mathbf{v} = \mathbf{u}$ and similarly the projection onto W along U is the mapping $P_w : V \to V$, defined by $P_w\mathbf{v} = \mathbf{w}$

Definition 4.7. (*Direct Sum in n dimensions*) Let $U_1, \dots, U_m$ be subspaces of V. We say that the sum $U_1 + \cdots + U_m$ is a direct sum and we write $V = U_1 \oplus \cdots \oplus U_m$ if $V = U_1 + \cdots + U_m$ and

$$U_i \cap \sum_{j \neq i} U_j \neq \{\mathbf{0}_v\} \qquad \forall\, i \in [m]$$

or equivilently, if $\forall\, \mathbf{v} \in V$ may be expressed uniquely as a sum,

$$\mathbf{v} = \mathbf{u}_1 + \cdots + \mathbf{u}_m \qquad \mathbf{u}_i \in U$$

Notation. I denote the set $[1, m] \cap \mathbb{Z}$ as $[m]$, for ease of LaTeX

5 Linear Transformations

Definition 5.1. (*Linear Transformation*) Let $(V, \oplus, \odot)$ and $(W, \boxplus, \boxdot)$ be vector spaces over $\mathbb{F}$. A function $T : V \to W$ is said to be a linear transformation if,

(i) $T(\mathbf{v}_1 \oplus \mathbf{v}_2) = T(\mathbf{v}_1) \boxplus T(\mathbf{v}_2)$

(ii) $T(a \odot \mathbf{v}) = a \boxdot T(\mathbf{v})$

If T is a linear transformation, $T : V \to V$, then we call it a linear operator.

Let $T : V \to W$ be a linear transformation and $P = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ be a basis for V. Say we know $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n) \in W$. Now let $\mathbf{v} \in V$ be an arbitrary element. Since P is a basis we can write, $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ for unique a_i,

$$\begin{aligned} T(\mathbf{v}) &= T\left(\sum_{i=1}^{n} a_i \mathbf{v}_i\right) \\ &= \sum_{i=1}^{n} a_i T(\mathbf{v}_i) \end{aligned}$$

So to find the actions on $\mathbf{v} \in V$ we simply take the linear combination of the vectors $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ with some scalars $a_1, \dots, a_n$.

Let $T : V \to W$ where $\mathbb{R}^n$ and $W = \mathbb{R}^m$. Let $\mathbf{v} = (a_1, \dots, a_n)$, then, in the standard basis we have,

$$\begin{aligned} T(\mathbf{v}) &= T\left(\sum_{i=1}^{n} a_i \mathbf{e}_i\right) \\ &= \sum_{i=1}^{n} a_i T(\mathbf{e}_i) \\ &= \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \end{aligned}$$

The matrix

$$\begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{pmatrix}$$

is called the Matrix of Transformation.

Theorem 5.2. Let V and W be finite vector spaces over $\mathbb{F}$ with a basis, $P = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ and $Q = [\mathbf{w}_1, \dots, \mathbf{w}_n]$ and let $T : V \to W$ be a linear transformation. Then $\exists$ a unique $[T]_P^Q \in M_{m\times n}(\mathbb{F})$, $\forall \mathbf{v} \in V$,

$$[T(\mathbf{v})]_Q = [T]_P^Q [\mathbf{v}]_P$$

Furthermore,

$$[T]_P^Q = \begin{pmatrix} [T(\mathbf{v}_1)]_Q & [T(\mathbf{v}_2)]_Q & \dots & [T(\mathbf{v}_n)]_Q \end{pmatrix}$$

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "{[\cdot]_P}"'] & W \arrow[d, "{[\cdot]_Q}"] \\ \mathbb{F}^n \arrow[r, "{[T]_P^Q}"] & \mathbb{F}^m \end{tikzcd}$$

Lemma 5.3. Let U, V and W be finite vector spaces over $\mathbb{F}$ and let $T : U \to V$ and $S : V \to W$ be linear transformations. Then, the composition $S \circ T : U \to W$ is also a linear transformation.

Theorem 5.4. Let U, V and W be finite vector spaces over $\mathbb{F}$ with ordered bases P, Q, R. Let $T : U \to V$ and $S : V \to W$ be linear transformations. Then, the matrix of $S \circ T$ wrt P and R is,

$$[S \circ T]_P^Q = [S]_Q^R [T]_P^Q$$

Definition 5.5. (*Idenity operator*) If V has bases P and P', so we define $\mathrm{id}_V : V \to V$ by $\mathrm{id}_V(\mathbf{v}) = \mathbf{v}$

Definition 5.6. (*Change of Basis Matrix*) Let V be a finite vector space a basis P, P'. Then $[id_V]_P^{P'}$ of the identity operator wrt, P, P' the change of basis matrix.

Corollary 5.7. With definition 5.6, $[id_V]_P^{P'}$ is invertible. Moreover,

$$\left([id_V]_P^{P'}\right)^{-1} = [id_V]_P^{P'}$$

Corollary 5.8. V has a basis P, P' and W is a finite vector space with a basis Q, Q'. Suppose $T : V \to W$ is a linear transformation. Then given $[T]_P^Q$ of T with respect to P, Q. Then, T with respect to P' and Q' is,

$$[T]_{P'}^{Q'} = [id_W]_Q^{Q'} [T]_P^Q [id_V]_{P'}^P$$

Definition 5.9. (*Similar Matrices*) Let A, $B \in M_{n\times n}(\mathbb{R})$. We say that B is similar to A if $\exists\, C \in M_{n\times n}(\mathbb{R})$, st, $B = CAC^{-1}$.

Corollary 5.10. If B is similar to A, A is similar C,

$$\begin{aligned} B &= CAC^{-1} \\ C^{-1}B &= CA \\ C^{-1}BC &= A \end{aligned}$$

Similarity is also really just a change of basis in the form,

$$\begin{aligned} CAC^{-1} &= B \\ [id_V]_P^Q [T]_P^P [id_V]_Q^P &= [T]_Q^Q \end{aligned}$$

Definition 5.11. (*Kernel*) Let $T : V \to W$ be a linear transformation. We define the Kernel of T (aka. null space) by,

$$\mathrm{Ker}(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}_w\}$$

Definition 5.12. (*Range*) We define $\mathrm{Range}\, T$ (aka. image) by,

$$\mathrm{Range}(T) = \{T(\mathbf{v}) = \mathbf{v} \in V\}$$

Lemma 5.13. Let $A \in M_{n\times n}(\mathbb{F})$, let T_A denote the associated linear transformation. Let A_{rref} be the rref of A and let $r \in \mathbb{N}$ be the number of pivots in A_{rref}. With ref the solution of $A\mathbf{x} = \mathbf{0}_{\mathbb{F}^m}$ with $\mathbf{x} = (x_1, \ldots, x_n)$, let $x_1, \ldots, x_{i_r}$ be pivots and $x_{i_{r+1}}, \ldots, x_{i_n}$ be the free variables.

(i) $\forall\, n - r$, free variables one to one and others to 0 and calculate $x_i, \ldots, x_{i_r}$. The resulting $\mathbf{x}$ form a basis for $\mathrm{Ker}\, A$

(ii) The r columns of A corresponding to pivot columns of A_{rref} form a basis for $Range(A)$

Lemma 5.14. A linear transformation $T : V \to W$ is injective if and only if $\mathrm{Ker}\ T = \mathbf{0}_v$

Proof. Suppose that $\operatorname{Ker} T = \{\mathbf{0}_v\}$. If $T(\mathbf{v}) = T(\mathbf{v}')$, then,

$$\begin{aligned} T(\mathbf{v} - \mathbf{v}') &= T(\mathbf{v}) - T(\mathbf{v}') \\ &= 0 & \text{Hence, } \mathbf{v} - \mathbf{v}' \in \operatorname{Ker} T \end{aligned}$$

Hence, $\mathbf{v} = \mathbf{v}'$. Hence, T is injective.
Conversely assume T is injective., Since $T(\mathbf{0}_v) = \mathbf{0}_w$, we have $\mathbf{0}_v \in \operatorname{Ker} T$. Since T is injective $\mathbf{0}_v$ must be the only vector in V that maps to $\mathbf{0}_w$. Hence $\operatorname{Ker} T = \{\mathbf{0}_w\}$ □

Definition 5.15. (*Rank*) The rank of T is the dimension of $\operatorname{Range} T$

Definition 5.16. (*Nullity*) The nullity is the dimension of $\operatorname{Ker} T$

Remark. For $A \in M_{n\times n}(\mathbb{F})$ the rank is equal to the number of pivots in A_{rref}, while nullity is equal to the number of free variables in A_{rref} or equivilently the number of columns minus the rank.

Corollary 5.17. Let $T : V \to W$ be a linear transformation. Then the following are equivalent:

(i) T is injective

(ii) $\operatorname{Ker} T = \{\mathbf{0}_v\}$

(iii) $\operatorname{null} T = 0$

If W is finite dimensional, then the following are equivilent,

(i) T is surjective

(ii) $\operatorname{Range} T = W$

(iii) $\operatorname{rank} T = \dim W$

Theorem 5.18. (*Rank Nullity Theorem*) Let V and W be finite vector spaces and let $T : V \to W$ be a linear transformation, then,

$$\dim V = \operatorname{rank} T + \operatorname{null} T$$

Corollary 5.19. Let $T : V \to W$ be a linear transformation from $n = \dim V \to m = \dim W$, then,

(i) $\operatorname{rank} T \leq \min\{m, n\}$

(ii) T is injective $\iff \operatorname{rank} T = n$

(iii) T is surjective $\iff \operatorname{null} T = n - m$

(iv) If $n > m$, then T is not injective

(v) If $n < m$, then T is not surjective

(vi) If $n = m$, then T is bijective

Lemma 5.20. Let $T : V \to W$ be a linear transformation. If T is a bijection, then its inverse $T^{-1} : W \to V$ is also a bijection linear transformation.

Definition 5.21. Let V and W be finite vector spaces over $\mathbb{F}$. If $\exists$ a bijective linear transformation $T : V \to W$, then T is said to be an isomorphism and the spaces V and W are said to be isomorphic.

Corollary 5.22. If V and W are isomorphic, W and V are isomorphic.

Proof. The proof follows nicely from Lemma 5.20 □

Theorem 5.23. Let V and W be finite vector spaces over $\mathbb{F}$. If V and W have the same dimension, then V and W are isomorphic.

Proof. Let $P = [\mathbf{v}_1, \ldots, \mathbf{v}_d]$ and $Q = [\mathbf{w}_1, \ldots, \mathbf{q}_d]$ be ordered basis of V and W. Now define $T : V \to W$ by the action on P as, $T(\mathbf{v}_i) = w_i \quad \forall\, i \in [d]$. By lemma now show that T is a bijection. $\square$

Definition 5.24. Let V be a vector space. Then a linear operator $P : V \to V$, st, $P^2 = P$ is called a projection operator.

Theorem 5.25. Let V be a vector space and U and W be subspaces of V. Let $P : V \to V$, then the following are equivilent,

(i) $V = U \oplus W$ and P is the projection map onto U along W.

(ii) P is a projection operator with $\operatorname{Range} P = U$ and $\operatorname{Ker} P = W$.

Moreover $Q = \mathrm{id}_V - P$ is the projection operator onto W along U.

6 Eigenvectors and Eigenvalues

Definition 6.1. Let V be a vectorspace over a field $\mathbb{F}$ and let $T : V \to V$ be a linear operator. A non-zero vector $\mathbf{v} \in V$ is said to be an eigenvector of T if $\exists\, \lambda \in \mathbb{F}$, st, $T(\mathbf{v}) = \lambda\mathbf{v}$

Remark. Let V be a vector space, $T : V \to V$ be a linear operator, and $\mathrm{id}_V : V \to V$ be the identity transformation. We then have,

$$\begin{aligned} T(\mathbf{v}) = \lambda\mathbf{v} \iff \mathbf{0}_v &= T(\mathbf{v}) - \lambda\mathbf{v} \\ &= T(\mathbf{v}) - \lambda\,\mathrm{id}_V(\mathbf{v}) \\ &= (T - \lambda\,\mathrm{id}_V)(\mathbf{v}) \end{aligned}$$

That is, $\mathbf{v}$ is an eigenvector of T with λ if and only if $\mathbf{v} \in \mathrm{Ker}(T - id_V) \setminus \{\mathbf{0}_v\}$. Also note if λ is an eigenvalue, $T - \lambda\,\mathrm{id}_V$ is not invertible.

Corollary 6.2. Let V be a vector space and $T : V \to V$ be a linear operator. Let λ be an eigenvalue of T. Then the set,

$$E_\lambda = \{\mathbf{v} \in V : T(\mathbf{v}) = \lambda\mathbf{v}\}$$

is a subspace of V, called the eigenspace of T associated to λ.

Proof. By the previous remark, $E_\lambda = \mathrm{Ker}(T - \lambda\,\mathrm{id}_V)$, ie. the kernel of the linear operator $T - \lambda\,\mathrm{id}_V$ and hence a subspace of V. □

Example. Let $A = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}$ and we have eigenvectors of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with $\lambda = 3$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ with $\lambda = 5$. We can also find eigenspaces,

$$\begin{aligned} E_3 = \mathrm{Ker}(A - 3I) &= \mathrm{Ker}\left(\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}\right) = \mathrm{span}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ E_5 = \mathrm{Ker}(A - 5I) &= \mathrm{Ker}\left(\begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}\right) = \mathrm{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \end{aligned}$$

Consider $D : C^\infty \to C^\infty$, what are the eigenvectors and λ's? We need to find f such that, $Df = \lambda f$, which is just a differential equation. Hence, the eigenvectors are, $f = Ae^{\lambda x} \quad (A \in \mathbb{R})$. So what about the eigenspaces? Then, $E_\lambda = \{x \mapsto ae^{\lambda x} : a \in \mathbb{R}\}$ so there are infinte eigenvectors and infinite λ's.

Theorem 6.3. Let V be a finite dimensional vector space with ordered basis P, and let T be a linear operator on V. Then $\mathbf{v} \in V$ is an eigenvectors of T if and only if $[\mathbf{v}]_P$ is an eigenvector of $[T]_P^P$; furthermore, the corresponding eigenvalues are the same.

Lemma 6.4. Let $\mathbb{F}$ be a field and let A be a square matrix over $\mathbb{F}$. Then A is invertible if and only if $\det A \neq 0$

Lemma 6.5. Let A be an $n \times n$ matrix, then λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$

Definition 6.6. Let $A \in M_{n\times n}(\mathbb{F})$, it's characteristic polynomial, p_A is defined by, $p_A(x) = \det(A - \lambda I)$

Theorem 6.7. The eigenvalues of a matrix A are precisely the roots of the characteristic equation.

6.1 Similar Matrices

Recall that two matrices A and B are called similar if there exists an invertible matrix C, such that,

$$B = CAC^{-1}$$

Similar matrices represent the same linear transformation just written with respect to two different bases. Hence, they must have the same eigenvalues since they have the same λ's as the transformation.

Theorem 6.8. Let $A \in M_{n\times n}$, let C be invertible $\in M_{n\times n}$ and $B = CAC^{-1}$. Then $\mathbf{v}$ is an eigenvector of A with λ iff $C\mathbf{v}$ is an eigenvector of C with λ.

Proof. If $A\mathbf{v} = \lambda\mathbf{v}$, then:
$$B(C\mathbf{v}) = CAC^{-1}(C\mathbf{v}) = CA\mathbf{v} = C\lambda\mathbf{v} = \lambda(C\mathbf{v})$$
Conversely, if $B(C\mathbf{v}) = \lambda C\mathbf{v}$, then,
$$A\mathbf{v} = C^{-1}CAC^{-1}C\mathbf{v} = C^{-1}BC\mathbf{v} = C^{-1}\lambda C\mathbf{v} = \lambda\mathbf{v}$$
□

Lemma 6.9. Let A and B be similar. Then $p_A = p_B$

Proof. Let $B = CAC^{-1}$, recall that $P_B = \det xI - B$. We have,
$$xI - B = xI - CAC^{-1} = xCIC^{-1} - CAC^{-1} = C(xI - A)C^{-1}$$
so,
$$\begin{aligned} p_B &= \det C(xI - A)C^{-1} \\ &= \det C \det xI - C \det C^{-1} \\ &= \det xI - A \det C \det C^{-1} \\ &- \det xI - A \det 1 \\ &= p_A \end{aligned}$$
□

Definition 6.10. (*Characteristic Polynomial wrt Ordered Basis*) Let V be a finite vector space and $T : V \to V$ be a linear transformation. Fix an arbitrary ordered basis P of V. The characteristic polynomial of p_T of T is p_A where $A = [T]_P^P$.

Definition 6.11. Let V be a vector space over a field $\mathbb{F}$ and let $T : V \to V$ be a linear operator. Let $f \in p(\mathbb{F})$ be a polynomial with polynomials with coefficients in $\mathbb{F}$, say,
$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$$
We define $f(T)$ to be the linear operator,
$$f(T) = a_n T^n + a_{n-1}T^{n-1} + \cdots + a_0$$
where $t^j = T^{j-1} \circ T$.

Theorem 6.12. (*Cayley Hamilton Theorem*) Let V be a finite vector space and let $T : V \to V$ be a linear transformation. If p_T is the characteristic polynomials of T, then $p_T(T) = 0_{L(V,V)}$.

6.2 Diagonisation

Here we assume,

- V is a finite vectorspace.
- $T : V \to V$ is linear operator.
- We have a basis of eigenvectors, $P = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ with $\lambda_1, \ldots \lambda_n$.

We can write an arbitrary $\mathbf{v} \in V$ as,

$$\mathbf{v} = a_1\mathbf{v}_1 + \dots a_n\mathbf{v}_n$$

for some $a_i \in \mathbb{F}$. Then,

$$\begin{aligned} T(\mathbf{v}) &= T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) \\ &= a_1T(\mathbf{v}_1) + \cdots + a_nT(\mathbf{v}_n) \\ \implies [T(\mathbf{v})]_P &= \begin{pmatrix} \lambda_1 a_1 \\ \vdots \\ \lambda_n a_n \end{pmatrix} \end{aligned}$$

and so it nicely follows that,

$$\begin{aligned} [T]_P^P &= \begin{pmatrix} [T(\mathbf{v}_1)]_P & \dots & [T(\mathbf{v}_n)]_P \end{pmatrix} \\ &= \begin{pmatrix} [\lambda_1\mathbf{v}_1]_P & \dots & [\lambda_n\mathbf{v}_n]_P \end{pmatrix} \\ &= \begin{pmatrix} \lambda & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \end{aligned}$$

Definition 6.13. (*Diagonalisable*) Let $T : V \to V$ be a linear operator. Then T is said to be diagonalisable if V has a basis consisting of eigenvectors of T.

Lemma 6.14. Let V be a vectorspace and let $T : V \to V$ be a linear operator. Suppose that S is a list of eigenvectors of T such that, no two elements of S are associated with the same eigenvalue, then S is linearly independent.

Proof. Exercise □

Corollary 6.15. Let V be a vectorspace of finite dimension n. If $T : V \to V$ has n distinct eigenvalues, then T is diagonalisable.

Remark. Take $\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$ we have two eigenvectors $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ but they are both associated with $\lambda = 3$ and hence you don't need to have distinct eigenvalues for a linear transformation.

6.3 Diagonalisation and Direct Sums

Corollary 6.16. Let V be a finite vectorspace of dimension n. Let $T : V \to V$ be a linear operator. Denote the distinct eigenvalues of T, $\lambda_1, \dots, \lambda_m$ ($m \leq n$). Then T is diagonal if and only if, $V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_m}$

Corollary 6.17. Let V be a vectorspace. Let $T : V \to V$ be a linear transformation with $\lambda_1, \dots, \lambda_n$. Then T is diagonalisable if and only if,

$$\dim V = \dim E_{\lambda_1} + \cdots + \dim E_{\lambda_n}$$

Assume m-dimensional vectorspace V with a diagonalisable linear operator $T : V \to V$, basis of eigenvectors $P = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ and eigenvalues l_i. Now applying T to $\mathbf{v}_i$, we get,

$$\begin{aligned} T(\mathbf{v}_i) &= \lambda_i\mathbf{v}_i \\ T^2(\mathbf{v}_i) &= \lambda_i^2\mathbf{v}_i \\ &\vdots \\ T^n(\mathbf{v}_i) &= \lambda_i^n\mathbf{v}_i \end{aligned}$$

Now, take $\mathbf{v} \in V$, $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m$ and so,

$$T^n(\mathbf{v}) = \lambda_1^n a_1 \mathbf{v}_1 + \cdots + \lambda_m^n a_m \mathbf{v}_n$$

Hence, we can now write that,

$$[T^n]_P^P = ([T]_P^P)^n = \begin{pmatrix} \lambda_1^n \ldots 0 & & \\ \vdots & \ddots & \vdots \\ 0 & \ldots \lambda_m^n & \end{pmatrix}$$

Now, letting $A \in M_{n\times n}(\mathbb{F})$ and B is a similar matrix to A.

$$\begin{aligned} A &= CBC^{-1} \\ A^2 &= CBC^{-1}CBC^{-1} = CB^2C^{-1} \\ &\vdots \\ A^n &= CB^nC^{-1} \end{aligned}$$

If A is diagonalisable, $A = CDC^{-1}$ with D a diagonalisable matrix,

$$A^n = CD^nC^{-1}$$

6.3.1 Linear Difference Equations

Consider a linear difference equation,

$$\mathbf{x}(n) = A\mathbf{x}(n-1) \quad \mathbf{x} : \mathbb{N}_0 \to \mathbb{F}^m \quad A \in M_{m\times m}(\mathbb{F})$$

This is generally coupled,

$$\begin{aligned} x_1(n) &= A_{11}x_1(n-1) + A_{12}x_2(n-1) + \cdots + A_{1m}x_m(n) \\ &\vdots \\ x_m(n) &= A_{1m}x_1(n-1) + A_{m2}x_2(n-1) + \cdots + A_{mm}x_m(n) \end{aligned}$$

However, if we know the state at $x(0)$, then we can do the following,

$$\begin{aligned} \mathbf{x}(1) &= A\mathbf{x}(0) \\ \mathbf{x}(2) &= A\mathbf{x}(1) = A^n\mathbf{x}(0) \\ &\vdots \\ \mathbf{x}(n) &= A^n\mathbf{x}(0) \end{aligned}$$

Hence, $A^n = CD^nC^{-1}$, and so $\mathbf{x}(n) = CD^nC^{-1}\mathbf{x}_0$

$$\begin{tikzcd} {[\mathbf{x}(0)]_P} \arrow[r, "D^n"] \arrow[d, "C^{-1}"'] & {[\mathbf{x}(n)]_P} \arrow[d, "C"] \\ \mathbf{x}(0) \arrow[r, "A^n"] & \mathbf{x}(n) \end{tikzcd}$$

6.3.2 Exponentials of Diagonalisable Linear Operators

Let $T : V \to V$ where $V = \mathbb{R}^m$ or $V = \mathbb{C}^m$ and their associated $m \times m$ matrices. Recall, $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ and hence define $e^T = \sum_{n=0}^{\infty} \frac{T^n}{n!}$ which itself is a linear operator on V. Furthermore, let $A \in M_{n\times n}(\mathbb{F})$, then $e^A = \sum_{n=0}^{\infty} \frac{e^A}{n!}$.

Let $V = \mathbb{F}^m$, $T : V \to V$ be diagonalisable and $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ a basis of eigenvectors with λ_i's. Now,

$$\begin{aligned} e^T(\mathbf{v}_i) &= \sum_{n=0}^{\infty} \frac{T^n(\mathbf{v}_i)}{n!} \\ &= \sum_{n=0}^{\infty} \frac{\lambda_i \mathbf{v}_i}{n!} \\ &= \mathbf{v}_i \sum_{n=0}^{\infty} \mathbf{v}_i e^{\lambda_i} \end{aligned}$$

Hence e^T has the same eigenvectors as T but with e^{λ_i}. Now we can write,

$$[e^T]_P^P = \begin{pmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_m} \end{pmatrix}$$

6.4 Applications to Differential Equations

Consider,

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x}$$

Then, we can solve this as,

$$\mathbf{x}(t) = e^{At}\mathbf{x}_0 \quad \forall t \in \mathbb{R} \text{ and } \mathbf{x}(0) = \mathbf{x}_0$$

Now suppose that A is diagonalisable, then we can write that,

$$\mathbf{x}(t) = Ce^{Jt}C^{-1}\mathbf{x}_0$$

or we could work in the P basis,

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= A\mathbf{x}(t) \\ &= CDC^{-1}\mathbf{x}_0 \\ C^{-1}\frac{d\mathbf{x}}{dt} &= DC^{-1}\mathbf{x}_0 \\ \frac{d}{dt}[\mathbf{x}(t)]_P &= D[\mathbf{x}(t)]_P \\ [\mathbf{x}(t)]_P &= e^{Dt}[\mathbf{x}_0]_P & \text{Hence } a_i(t) = e^{\lambda_i t}a_i(0) \\ CC^{-1}[\mathbf{x}(t)]_P &= Ce^{Dt}C^{-1}[\mathbf{x}_0]_P \end{aligned}$$

$$\begin{tikzcd} \mathbf{x}(0) \arrow[r, "e^{At}"] \arrow[d, "C^{-1}"'] & \mathbf{x}(t) \\ {[\mathbf{x}(0)]_P} \arrow[r, "e^{Dt}"] & {[\mathbf{x}(t)]_P} \arrow[u, "C"'] \end{tikzcd}$$

6.5 Complex Eigenvalues

Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ be a real operator on $\mathbb{R}^2$ which has the p_A of; $p_A = x^2 + 1$, which has no real roots, hence A cannot have any eigenvalues such that $(\lambda \notin \mathbb{R})$. This can be done quickly by noting that every vector is rotated by 90° (no stretching). However, p_A has eigenvalues of $\pm i$. Instead, let $A \in M_{2\times 2}(\mathbb{C})$ instead of $M_{2\times 2}(\mathbb{R})$ and now we have eigenvectors,

$$\mathbf{v}_{1,2} = \left\{ \begin{pmatrix} 1 \\ -i \end{pmatrix}, \begin{pmatrix} 1 \\ i \end{pmatrix} \right\}$$

and we can say that,

$$A = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$$

which will turn out to be real. The expansion from $\mathbb{R}^2 \to \mathbb{C}^2$ is natural and useful as there are more tools. More generally, if V is a real vector space, then we take the complexification $V^{\mathbb{C}}$, to be the vectorspace where we now take complex linear combinations of vectors in V. Complexification is defined as the tensor product of V with $\mathbb{C}$.

$V \subseteq V^{\mathbb{C}}$. Considering the complexification of a space makes no difference to the action of a linear operator on V, but we can just use complex eigenvectors. All of this basically boils down to the fact that $\mathbb{R}$ isn't an algebraically closed field

7 Jordan Normal Form

Example. Let $A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$ and we can see that A isn't diagonalisable. There isn't a basis of eigenvectors, however it's almost diagonalisable, theres only a single one on the super diagonal. Now consider the affect of A on the standard basis,

$$\begin{aligned} A\mathbf{e}_1 &= 4\mathbf{e}_1 &&\implies A^n\mathbf{e}_1 = 4^n\mathbf{e}_1 \\ A\mathbf{e}_2 &= 3\mathbf{e}_2 &&\implies A^n\mathbf{e}_2 = 3^n\mathbf{e}_2 \\ A\mathbf{e}_3 &= \mathbf{e}_2 + 3\mathbf{e}_3 &&\implies \quad ? \end{aligned}$$

So we look at $A^2\mathbf{e}_3$ and try and find a pattern,

$$\begin{aligned} A^2\mathbf{e}_3 &= A(\mathbf{e}_2 + 3\mathbf{e}_3) \\ &= A\mathbf{e}_2 + 3A\mathbf{e}_3 \\ &= 2 \cdot 3\mathbf{e}_2 + 3^2 \cdot \mathbf{e}_3 \end{aligned}$$

and now, to confirm our suspicions, we try $A^3\mathbf{e}_3$,

$$\begin{aligned} A^3\mathbf{e}_3 &= A(2 \cdot 3\mathbf{e}_2 + 3^2 \cdot \mathbf{e}_3) \\ &= 2 \cdot 3A\mathbf{e}_2 + 3^2\mathbf{e}_3 \\ &= 3 \cdot 3^2\mathbf{e}_2 + 3^n\mathbf{e}_3 \end{aligned}$$

which confirms it and so we say, $A^n\mathbf{e}_3 = n3^{n-1}\mathbf{e}_2 + 3^n\mathbf{e}_3$. So we can write,

$$A^n = \begin{pmatrix} 4^n & 0 & 0 \\ 0 & 3^n & n3^{n-1} \\ 0 & 0 & 3^n \end{pmatrix}$$

Definition 7.1. (*Jordan Block Matrix*) A $n \times n$ matrix that has diagonal entries λ and 1's on the superdiagonal and zeros elsewhere i.e.

$$\begin{pmatrix} \lambda & 1 & \dots & 0 & 0 \\ 0 & \lambda & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

Note: Any 1×1 matrix is considered a Jordan Block matrix even though there is no superdiagonal.

Definition 7.2. (*Jordan Block Basis*) Let V be a finite vector space $P = [\mathbf{v}_1, \dots, \mathbf{v}_d]$. Let $T : V \to V$ be a linear transformation. The basis P is said to be a jordan block matrix for T if $\exists\, \lambda \in \mathbb{R}$,

$$T(\mathbf{v}_j) = \begin{cases} \lambda\mathbf{v}_j & \text{if j = 1} \\ \mathbf{v}_{j-1} + \lambda\mathbf{v}_j & \text{otherwise} \end{cases}$$

Corollary 7.3. Let V be a finite vectorspace, $T : V \to V$ be a linear transformation and P be a basis of V. Then P is a jordan block basis if and only if $[T]_P^P$ is a jordan block matrix.

Proof. Trivial $\square$

Definition 7.4. (*Jordan Normal Form*) A matrix is in Jordan Form or Jordan Canonical Form if it is a block diagonal matrix in which of the blocks is a Jordan Block Matrix.

Block Diagonal Matrix: $\begin{pmatrix} A & & 0 \\ & B & \\ 0 & & C \end{pmatrix}$ $\begin{pmatrix} a & b & c & & & \\ d & e & f & & & \\ g & h & i & & & \\ & & & j & k & \\ & & & l & m & \\ & & & & & n \end{pmatrix}$

Theorem 7.5. Let V be a finite vector space and $T : V \to V$ be a linear operator whose characteristic polynomial splits into linear factors. Then there exists a basis of V with respect to which $[T]_P^P$ is in Jordan Normal Form. We call such a jordan basis. Moreover, the Jordan Normal Form is unique up to permutations of the jordan basis.

Similar Matrices have the same Jordan Normal Form.

Corollary 7.6. Let A and B be $n \times n$ matrices each with a Jordan Normal Form. Then A and B are similar if and only if they both have the same Jordan Normal Form (up to permutations).

Theorem 7.7. (*FTA*) Every non-zero polynomial in $\mathbb{C}$ of degree n with complex coefficients has exactly n complex roots. (counted with multiplicity).

The upshot is that every polynomial in $\mathbb{C}$ splits into linear factors. So now we restrict ourselves to Complex Vector Spaces.

Definition 7.8. (*Minimum Polynomial*) Let $A \in M_{n\times n}(\mathbb{F})$. The minimum polynomial m_A of A is a monic polynomial of minimum positive degree such that $m_A(A) = \mathbf{0}$.

By Cayley Hamilton, we know that $\deg m_A \leq \deg p_A$.

Lemma 7.9. Let $A \in M_{n\times n}(\mathbb{F})$ with p_A, then,

- m_A is unique
- If f is a polynomial with coefficients in $\mathbb{F}$, st, $f(A) = \mathbf{0}$, then m_A divides f.
- A scalar λ is an eigenvalue of m_A if and only if $m_A(\lambda) = 0$. Hence, p_A and m_A have the same zeros.

Corollary 7.10. Let $A \in M_{n\times n}(\mathbb{F})$ with p_A and m_A. Suppose that,

$$p_A(x) = (x - \lambda_1)^{a_1} \dots (x - \lambda_d)^{a_d}$$

where $\lambda_1, \dots, \lambda_d$ are distinct eigenvalues of A. Then there exist some, $b_1, \dots, b_d$ such that $1 \leq b_j \leq a_j$ and,

$$m_A = (x - \lambda_1)^{b_1} \dots (x - \lambda_d)^{b_d}$$

Definition 7.11. Let $T : V \to V$ be a linear transformation on a finite vectorspace V. Suppose the characteristic polynomial of T is,

$$p_T = (x - \lambda_1)^{a_1} \dots (x - \lambda_d)^{a_d}$$

where $\lambda_1, \dots, \lambda_d$ are eigenvalues of T. Then a_i is called the algebraic multiplicity of λ_i. Then $g_i = \dim(E_{\lambda_i})$ of the eigenspace E_{λ_i} is called the multiplicity of λ_i.

Example. Let $A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$ and now find λ_i, a_i and g_i.
Firstly we can see that $p_A(x) = (x - 4)(x - 3)^2$,
$\lambda_1 = 4$, and so $a_1 = 1$ and $g_1 = \dim(E_4) = \dim \operatorname{span}([\mathbf{e}_1]) = 1$
$\lambda_2 = 3$ and so $a_2 = 2$ and $g_2 = \dim(E_3) = \dim \operatorname{span}([\mathbf{e}_2]) = 1$

Theorem 7.12. (*Jordan Normal Form Theorem*) Let T be a linear transformation and let A be its Jordan Normal Form. Let,

$$p_T(x) = (x-\lambda_1)^{a_1}\dots(x-\lambda_d)^{a_d} \text{ and } m_T = (x-\lambda_1)^{b_1}\dots(x-\lambda_d)^{b_d}$$

be the characteristic and minimum polynomials repsectively. Then,

- The arithmetic multiplicity a_i is the number of occurances of λ_i on the diagonal of A.
- The exponent b_i is the size of the largest λ_i-block in A.
- The geometric multiplicity g_i of λ_i is the number of λ_i blocks in A.

For a 3×3 matrix we only need the first two statements, and there are the following possibilities,

- $p_T = (x-\lambda_1)(x-\lambda_2)(x-\lambda_3)$ where $\lambda_1 \neq \lambda_2 \neq \lambda_3$

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

- $p_T = (x-\lambda_1)^2(x-\lambda_2)$ where $\lambda_1 \neq \lambda_2$

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix}$$

- $p_T = (x-\lambda)^3$

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix}$$

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

Example. $A = \begin{pmatrix} 5 & 4 & 0 \\ -1 & 0 & 1 \\ -1 & 5 & 6 \end{pmatrix}$

Now find the Jordan Normal Form. We see that $p_A = (x-5)(x-3)^2$ so try a $m_A = (x-5)(x-3)$,

$$m_A(A) = \begin{pmatrix} 0 & 4 & 0 \\ -1 & -5 & 1 \\ -1 & 5 & 1 \end{pmatrix}\begin{pmatrix} 2 & 4 & 0 \\ -1 & -3 & 1 \\ -1 & 5 & 3 \end{pmatrix} = \begin{pmatrix} -4 & * & * \\ * & * & * \\ * & * & * \end{pmatrix}$$

Hence $m_A(A) \neq 0$ and try $m_A = (x-5)(x-3)^2$, which does work and so $m_A = (x-5)(x-3)^2$. Hence we can say,

$$A_{JNF} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

7.1 Jordan Bases

Example. Taking A to be the same as the matrix in the last example, we can assume that we have $P = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ a jordan basis. Then, we apply A to each of these basis vectors,

$$\begin{aligned} A\mathbf{v}_1 &= 5\mathbf{v}_1 \\ A\mathbf{v}_2 &= 3\mathbf{v}_2 \\ A\mathbf{v}_3 &= \mathbf{v}_2 + 3\mathbf{v}_3 \end{aligned}$$

Hence we can say that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A with respect to eigenvalues of $\lambda = 5$ and $\lambda = 3$ respectively. However, $\mathbf{v}_3$ is not an eigenvector, but we can say the following,

$$\begin{aligned} (A - 3I)\mathbf{v}_3 &= A\mathbf{v}_3 - 3\mathbf{v}_3 \\ &= \mathbf{v}_2 \end{aligned}$$

and so we can say that,

$$\begin{aligned} (A - 3I)^2\mathbf{v}_3 &= (A - 3I)\mathbf{v}_2 \\ &= 0 \end{aligned}$$

so, $\mathbf{v}_3 \in \text{Ker}((A - 3I)^2)$, whilst, $\mathbf{v}_2 \in \text{Ker}(A - 3I)$. Now to find the vectors, the first two are quite easy to find as they are just linearly independent eigenvectors. $\mathbf{v}_3$ is slightly harder to find. Firstly compute $(A - 3I)^2$ and it's reduced row echelon form.

$$(A - 3I)^2 = \begin{pmatrix} 0 & -4 & 4 \\ 0 & 0 & 0 \\ 0 & -4 & 4 \end{pmatrix} \qquad (A - 3I)^2_{RREF} = \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and so now we can say,

$$\text{Ker}((A - 3I)^2) = \left\{ a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} : a, b \in \mathbb{C} \right\}$$

so we let $\mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

NB! $E_3 = \text{Ker}(A - 3I) \subseteq \text{Ker}((A - 3I)^2)$. When choosing $\mathbf{v}_3$ we don't want an eigenvector as what we want must be linearly independent.

Now define $\mathbf{v}_2 = (A - 3I)\mathbf{v}_3$ and so $(A - 3I)\mathbf{v}_2 = (A - 3I)^2\mathbf{v}_3 = \mathbf{0}$. i.e $\mathbf{v}_2 \in E_3 \setminus \{0\}$ as $\mathbf{v}_3$ is not an eigenvector. Moreover,

$$\begin{aligned} \mathbf{v}_2 &= (A - 3I)\mathbf{v}_3 \\ &= A\mathbf{v}_3 - 3\mathbf{v}_3 \\ &= \mathbf{v}_2 - 3\mathbf{v}_3 \end{aligned}$$

and so this satisfies the critrion for a Jordan Basis. Now, it becomes easy as we just choose our eigenvector and,

$$(A - 3I) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \qquad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

and we can check by letting $C = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{pmatrix}$ and so,

$$\mathcal{J} = C^{-1}AC = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

8 Generalised Eigenvectors

COMING SOON

9 Inner Product Spaces

The inner produce is the generalisation of dot product.

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$$

and

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Definition 9.1. Let V be a vector space. Then an inner product on V is a function which maps an ordered pair of vectors $(\mathbf{u},\, \mathbf{v}) \in V \times V$ to a number $\langle \mathbf{u} \,, \mathbf{v} \rangle \in \mathbb{F}$, and that satisfies the following properties,

(i) Linearity,

$$\langle a\mathbf{u} + b\mathbf{v} \,, \mathbf{w} \rangle = a \langle \mathbf{u} \,, \mathbf{w} \rangle + b \langle \mathbf{v} \,, \mathbf{w} \rangle$$

(ii) Conjugate Symmetry,

$$\langle \mathbf{u} \,, \mathbf{v} \rangle = \overline{\langle \mathbf{v} \,, \mathbf{u} \rangle}$$

(iii) Positivity, if $\mathbf{v} \neq 0$,

$$\langle \mathbf{v} \,, \mathbf{v} \rangle > 0$$

Such a vector space V equipped with an inner product is called an inner produce space.

Lemma 9.2. Let V be an inner product space. Then $\forall\, \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a,\, b \in \mathbb{F}$ then the following hold,

(i) Conjuage linearity,

$$\langle \mathbf{u} \,, a\mathbf{v} + b\mathbf{w} \rangle = \overline{a} \langle \mathbf{u} \,, \mathbf{v} \rangle + \overline{b} \langle \mathbf{u} \,, \mathbf{w} \rangle$$

(ii) $\langle \mathbf{0} \,, \mathbf{v} \rangle = \langle \mathbf{v} \,, \mathbf{0} \rangle = 0$

(iii) $\langle \mathbf{v} \,, \mathbf{v} \rangle = 0 \iff \mathbf{v} = 0$

(iv) If $\langle \mathbf{v} \,, \mathbf{u} \rangle = \langle \mathbf{v} \,, \mathbf{w} \rangle \, \forall\, v \in V$, then $\mathbf{u} = \mathbf{w}$.

Definition 9.3. Let V be an inner product space. Then we define the norm of $\mathbf{v} \in V$ by,

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v} \,, \mathbf{v} \rangle}$$

If $\|\mathbf{v}\| = 1$, then $\mathbf{v}$ is a unit vector.

Lemma 9.4. Let V be an inner product space over $\mathbb{F}$, then, $\forall\, \mathbf{v} \in V$ and $a \in \mathbb{F}$,

(i) $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$

(ii) $\|a\mathbf{v}\| = |a| \, \|\mathbf{v}\|$

Theorem 9.5. (Cauchy-Schwartz Inequality) Let V be a inner product space. Then $\forall\, \mathbf{u}, \mathbf{v} \in V$,

$$| \langle \mathbf{u} \,, \mathbf{v} \rangle | \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$$

with equality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent

Definition 9.6. Let V be an inner product space. The angle between two vectors $\mathbf{u}$ and $\mathbf{v}$ in V is the unique value $\theta \in [0,\, 2\pi]$ such that,

$$\cos\theta = \frac{\langle \mathbf{u} \,, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$$

Definition 9.7. (*Orthogonal*) Let V be an inner product space. Then two vectors $\mathbf{u}, \mathbf{v} \in V$ are said to be orthogonal if $\langle \mathbf{u} \,, \mathbf{v} \rangle = 0$. More generally, a set or list of vectors S of V is said to be orthogonal if any two vectors S are orthogonal.

Lemma 9.8. Let V be a inner product space and S be an orthogonal set or list of vectors from V. Then S is linearly independent.

9.1 Orthogonal Bases

Choosing the right bases can make things so much easier when talking about Linear Algebra. For an inner product space, we will see that the 'nicest' basis to work in is ones which contain mutually orthogonal vectors and have unit length. We shall start by considering a finite orthognal basis P (not necessarily unit length). Now $\forall\, \mathbf{v} \in V$, we can find $[\mathbf{v}]_P$ as follows using inner products,

Theorem 9.9. Let $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be an orthogonal basis for an inner product space V. Then $\forall\, \mathbf{v} \in V$,

$$\mathbf{v} = \sum_{i=1}^{n} \frac{\langle \mathbf{v} \,, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \mathbf{v}_i$$

and hence,

$$[\mathbf{v}]_P = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

with,

$$a_i = \frac{\langle \mathbf{v} \,, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|}$$

Proof. Since P is a basis,

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

for some $a_i \in \mathbb{F}$. Then,

$$\begin{aligned} \langle \mathbf{v} \,, \mathbf{v}_i \rangle &= \left\langle \sum_{j=1}^{n} a_j \mathbf{v}_j \,, \mathbf{v}_i \right\rangle \\ &= \sum_{j=0}^{n} a_j \langle \mathbf{v}_j \,, \mathbf{v}_i \rangle \\ &= a_i \langle \mathbf{v}_i \,, \mathbf{v}_i \rangle \\ &= a_i \|\mathbf{v}_i\|^2 \end{aligned}$$

and hence,

$$a_i = \frac{\langle \mathbf{v} \,, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2}$$

and hence,

$$\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{v}_i = \sum_{i=1}^{n} \frac{\langle \mathbf{v} \,, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \mathbf{v}_i$$

□

Definition 9.10. (*Orthonormal*) Let V be an inner product space. A set A or list S of vectors from V is said to be orthonormal if S is orthogonal and every vector in S is a unit vector.

In particular, if S is an orthonormal list, then,

$$\langle \mathbf{v}_i \,, \mathbf{v}_j \rangle = \delta_{ij}$$

which then reduces the previous theorem to say that,

$$a_i = \langle \mathbf{v} \,, \mathbf{v}_i \rangle$$

Theorem 9.11. Let $T : V \to W$ be a linear transformation between finite dimensional inner product spaces V and W with inner products of $\langle \cdot\, , \cdot \rangle_V$ and $\langle \cdot\, , \cdot \rangle_W$ respectively. Let $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be an orthonormal basis of V and $Q = [\mathbf{w}_1, \ldots, \mathbf{w}_n]$ be an orthonormal basis for W. Let $A = [T]_P^Q$. Then,

$$A_{ij} = \langle T(\mathbf{v}_j)\, , \mathbf{w}_i \rangle$$

Proof. Since Q is a basis we can write the vector $T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \cdots + a_{mj}\mathbf{v}_m$ for a unique scalars, $a_{ij} \in \mathbb{F}$. Then,

$$[T(\mathbf{v}_j)]_Q = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

From Thm 7.9, we can now rewrite it as,

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} \langle T(\mathbf{v}_j)\, , \mathbf{w}_i \rangle_W \mathbf{w}_i$$

and hence,

$$A_{ij} = \langle T(\mathbf{v}_j)\, , \mathbf{w}_i \rangle_W$$

□

Remark. Nowhere in the proof do we use that P is an inner product space. So we can generalise the theorem further with P not being a .

9.2 Orthogonal Projections

Theorem 9.12. Let $\mathbf{v}$ and $\mathbf{u}$ be vectors in an inner product space V, $\mathbf{u} \neq \mathbf{0}$. Then $\mathbf{v}$ may be expressed as a sum,

$$\mathbf{v} = \mathbf{u}' + \mathbf{z}$$

where $\mathbf{u}'$ is a scalar multiple $\mathbf{u}$ and $\mathbf{z}$ is orthogonal to $\mathbf{u}$. Furthermore,

$$\mathbf{u}' = \frac{\langle \mathbf{v}\, , \mathbf{u} \rangle}{\|\mathbf{u}\|^2}\mathbf{u}$$

We note here that this is a direct sum as we can represent every vector in V as the decomposition of two vectors, one from $\mathbf{u}'$ and another $\mathbf{z}$.

Proof. First assume we can write,

$$\mathbf{v} = \mathbf{u}' + \mathbf{z}$$

with $\mathbf{u}' = a\mathbf{u}$. Then,

$$\begin{aligned} \langle \mathbf{v}\, , \mathbf{u} \rangle &= \langle \mathbf{u}' + \mathbf{z}\, , \mathbf{u} \rangle \\ &= \langle a\mathbf{u} + \mathbf{z}\, , \mathbf{u} \rangle \\ &= a \langle \mathbf{u}\, , \mathbf{u} \rangle + \langle \mathbf{u}\, , \mathbf{z} \rangle \\ &= a \|\mathbf{u}\|^2 + 0 \end{aligned}$$

Hence,

$$a = \mathbf{v} - \frac{\langle \mathbf{v}\, , \mathbf{u} \rangle}{\|\mathbf{u}\|^2}$$

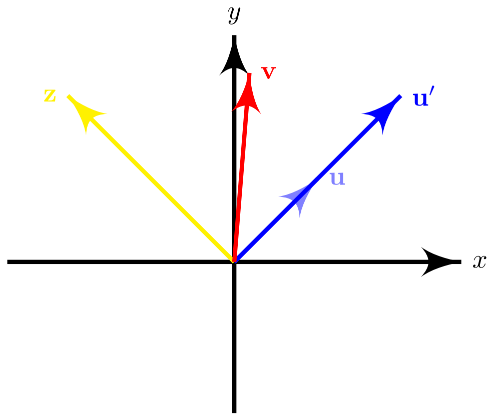

Figure 1: Orthogonal Projections Diagram

Now we check that,

$$\mathbf{z} = \frac{\langle \mathbf{v}\,, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\mathbf{u}$$

is orthogonal to $\mathbf{u}$,

$$\begin{aligned}
\langle \mathbf{z}\,, \mathbf{u}\rangle &= \left\langle \mathbf{v} - \frac{\langle \mathbf{v}\,, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\,, \mathbf{u}\right\rangle \\
&= \langle \mathbf{v}\,, \mathbf{u}\rangle - -\frac{\langle \mathbf{v}\,, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\langle \mathbf{u}\,, \mathbf{u}\rangle \\
&= \langle \mathbf{v}\,, \mathbf{u}\rangle - \langle \mathbf{v}\,, \mathbf{u}\rangle \\
&= 0
\end{aligned}$$

□

Definition 9.13. With the notation of the above theorem, the linear operator $P_{\mathbf{u}} : V \to V$ is defined by,

$$P_{\mathbf{u}} = \frac{\langle \mathbf{v}\,, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\mathbf{u} = \mathbf{u}'$$

is called the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$

Theorem 9.14. Let $S = [\mathbf{v}_1, \ldots, \mathbf{v}_m]$ be an orthogonal list of non-zero vectors in an inner product space V. Let $\mathbf{v}$ be an arbitrary vector in V. Then $\mathbf{v}$ may be uniquely expressed as a sum,

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m + \mathbf{z}$$

with $\mathbf{z}$ orthogonal to every vector in S. Furthermore,

$$a_i\mathbf{v}_i = P_{\mathbf{v}_i}(\mathbf{v})$$

Remark. This implies that we can write a $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ as $\mathbf{v} = \sum_{i=1}^{n} P_{\mathbf{v}_i}(\mathbf{v})$

9.3 The Gram-Schmidt Process

Let $Q = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ be an arbitrary basis of an inner product space V. We first construct an orthognal basis $R = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ from Q using the preceeding theorem.

– Let $\mathbf{v}_1 = \mathbf{u}_1$

– For $k > 1$, let,

$$\begin{aligned}\mathbf{v}_k &= \mathbf{u}_k - \sum_{i=1}^{k-1} P_{\mathbf{v}_i}(\mathbf{u}_k) \\ &= \mathbf{u}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{u}_k \,, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \mathbf{v}_i\end{aligned}$$

– Normalise the vectors $\mathbf{v}_i$ to obtain $\mathbf{w}_i = \dfrac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$

Then $S = [\mathbf{w}_1, \ldots, \mathbf{w}_n]$ is our orthonormal basis.

Remark. This implies for every finite dimensional inner product spaces, there is always an orthonormal basis.

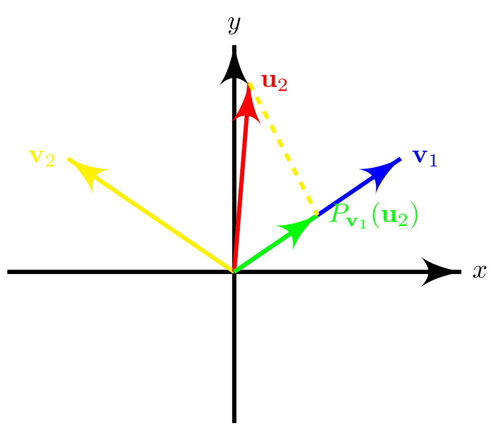

Figure 2: An orthognal basis for $\mathbb{R}^2$

Definition 9.15. Let V be an inner product space and let U be a subspace of V. Then the orthonormal complement of U, denoted $U^\perp$, is the set of vectors that are orthogonal to every vector of U. That is,

$$U^\perp = \{\mathbf{v} \in V : \langle \mathbf{v} \,, \mathbf{u} \rangle = 0 \forall\, \mathbf{u} \in U\}$$

Example. Let $U = \{\mathbf{x} \in \mathbb{R}^3 : x_1 = x_2 = 0\}$, then $U^\perp = \{\mathbf{x} \in \mathbb{R}^3 : x_1 = 0\}$.

One can also show that $U^\perp$ is a subspace.

Theorem 9.16. Let V be an inner product space and let U be a finite dimensional subspace of V. Then, $V = U \oplus U^\perp$

Proof. Let $\mathbf{v} \in V$ and $\mathbf{u} \in U$. Then we know by thm13.19 that we can write,

$$\mathbf{v} = \mathbf{u}' + \mathbf{z}$$

with $\mathbf{u}' = a\mathbf{u} \in U$, and $z \in U^\perp$. Hence any vector $\mathbf{v} \in U$ is in $U + U^\perp$. Hence $V \subset U + U^\perp$. Clearly $U + U^\perp \subset V$. Therefore,

$$V = U + U^\perp$$

So now take $\mathbf{v} \in U \cap U^\perp$. Then, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ and by lemma 13.5, this implies that, $\mathbf{v} = 0$. Hence, $U \cap U^\perp = \{\mathbf{0}\}$, and

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

□

Definition 9.17. (*Orthogonal Projection*) The projection P_U onto a subspace Ualong it's orthogonal complement $U^\perp$ is called the orthogonal projection onto U.

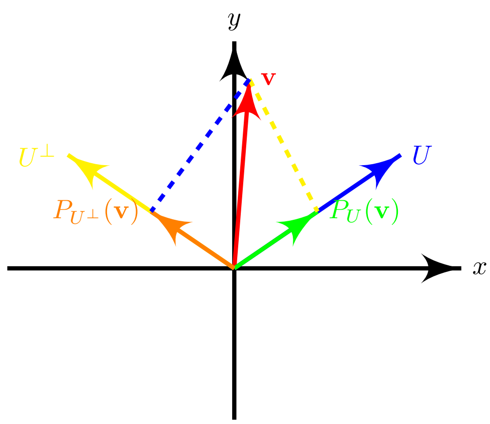

Figure 3: Orthogonal Projections, $\mathbf{v} = P_U(\mathbf{v}) + P_{U^\perp}(\mathbf{v})$

It's pretty simple to construct these operators using orthogonal projections onto vectors. Let $S = [\mathbf{u}_1, \ldots, \mathbf{v}_m]$ be an orthogonal basis of U. then for an arbitrary $\mathbf{v} \in V$ we have,

$$P_U(\mathbf{v}) = \sum_{i=1}^{m} P_{\mathbf{u}_i}(\mathbf{v}) = \sum_{i=1}^{m} \frac{\langle \mathbf{v}, \mathbf{u}_i \rangle}{\|\mathbf{u}_i\|^2} \mathbf{u}_i$$

and if S is orthonormal,

$$P_U(\mathbf{v}) = \sum_{i=1}^{m} \langle \mathbf{v}, \mathbf{u}_i \rangle \mathbf{u}_i$$

To see this formula works and gives all the required projections, note that $P_U(\mathbf{v}) \subset U$ and the vector $\mathbf{v} - P_U(\mathbf{v})$ is perpendicular to all of the $\mathbf{u}_i$'s hence is in $U^\perp$

9.4 Least Squares

We often have a system of equations without a solution and so we can use least squares to find a 'close enough' solution.

Let us say we have a load of points that we think fit a line $y = b_0 + b_1 x$. Then we have a set of n linear equations,

$$b_0 + b_1 x_1 = y_1$$
$$\vdots$$
$$b_0 + b_1 x_n = y_n$$

which we can write as,

$$A\mathbf{v} = \mathbf{w}$$

Typically there doesn't exist a solution to this, but we are going to look for a line that minimises the differences between the line and the actual y_i values. We will see that this can be just done through an orthogonal projection.

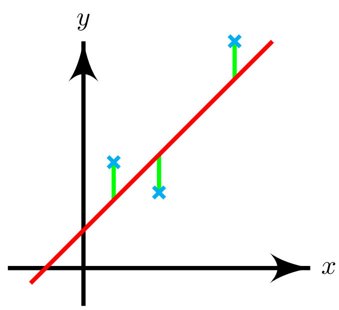

Figure 4: Least Squares on an line

We will consider general sets of real linear equations, $A\mathbf{v} = \mathbf{w}$ with matrices $A \in M_{m\times n}(\mathbb{R})$,

- Let $\mathbf{a}_i$ be the ith column of A.
- Let U denote the column space of A ($U = \text{span}([\mathbf{a}_1, \ldots, \mathbf{a}_n])$)
- There is a solution if and only if $\mathbf{w} \in U$
- The solution is unique if and only if a_i's are linearly independent.

If $vecw \notin U$, then there are no solutions. However, we can look for a vector $\mathbf{v}$ that minimises error, $\|\mathbf{w} - A\mathbf{v}\|$. Note that,

$$\|\mathbf{w} - A\mathbf{v}\| = \langle \mathbf{w} - A\mathbf{v}\,, \mathbf{w} - A\mathbf{v}\rangle = \sum_{i=1}^{m} (w_i - (A\mathbf{v})_i)^2$$

hence, least squares.

Definition 9.18. Let $A \in M_{n\times m}(\mathbb{R})$ and $\mathbf{w} \in \mathbb{R}^m$. A least squares solution to the linear system of equations $A\mathbf{v} = \mathbf{w}$ is a vector $\mathbf{v}$ that minimises $\|\mathbf{w} - A\mathbf{v}\|$, with the norm defined as the standard inner product on $\mathbb{R}^m$. The quantity $E = \|\mathbf{w} - A\mathbf{v}\|$ is the error.

Corollary 9.19. Let U be a subspace of a finite dimensional subspace W, and let $\mathbf{w} \in W$. By Thm 13.23 we can write a unique decomposition $\mathbf{w} = \mathbf{u} + \mathbf{z}$, with $\mathbf{u} \in U$ and $z \in U^{\perp}$. then $\mathbf{u}$ is the unique vector in U that is closest to $\mathbf{w}$, in the sense $\forall\, \mathbf{u}' \in U$ with $\mathbf{u}' \neq \mathbf{u}$, $\|\mathbf{w} - \mathbf{u}'\| > \|\mathbf{w} - \mathbf{u}\|$.

Proof. $\mathbf{u} - \mathbf{u}' \in U$ (since U is a subspace), hence,

$$\begin{aligned}
\|\mathbf{w} - \mathbf{u}'\|^2 &= \|\mathbf{u} + \mathbf{z} - \mathbf{u}'\|^2 \\
&= \|\mathbf{u} - \mathbf{u}' + \mathbf{z}\|^2 \\
&= \|\mathbf{u} - \mathbf{u}'\| + \left\|\mathbf{z}^2\right\| && \text{since } \mathbf{u} - \mathbf{u}' \text{ and } \mathbf{z} \text{ are orthogonal.} \\
&> \|\mathbf{z}\|^2 \\
&= \|\mathbf{w} - \mathbf{u}\|^2
\end{aligned}$$

□

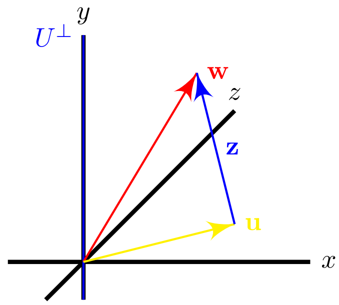

Figure 5: Diagram for Corollary 7.19

This result tells us for a vector $\mathbf{v}$ that minimises the error is a solution of,

$$A\mathbf{v} = \mathbf{u} \qquad \text{where } \mathbf{u} = P_U(\mathbf{w}), \text{ the orthogonal projection onto } U$$

Since $\mathbf{u} \in U$, the column space of A, there is indeed a solution. Moreover this solution is unique if and only if the columns of A are linearly independent. (hence, A is injective).

So how do we find $\mathbf{u}$? We need to find P_U. This is easy if the columns of A are orthogonal, however this isn't true for most of the time.

Theorem 9.20. The vector $\mathbf{v}$ is a least squares solution to $A\mathbf{v} = \mathbf{w}$ if and only if $\mathbf{v}$ is a solution of $A^T A\mathbf{v} = A^T\mathbf{w}$ if and only if it is the unique least squares solution to $A\mathbf{v} = \mathbf{w}$.

Proof. Let $\mathbf{u}$ be the orthogonal projection of $\mathbf{w}$ onto Range A. Then $\mathbf{z} = \mathbf{w} - \mathbf{u} \in (\text{Range } A)^{\perp} = \text{Ker } A^T$

Suppose $\mathbf{v}$ is a least squares solution to $A\mathbf{v} = \mathbf{w}$, i.e. $A\mathbf{v} = \mathbf{u}$, then,

$$\begin{aligned} A^T A\mathbf{v} &= A^T\mathbf{u} \\ &= A^T(\mathbf{w} - \mathbf{z}) \\ &= A^T\mathbf{w} && \text{as } \mathbf{z} \in \text{Ker } A^T \end{aligned}$$

This proves one direction. Next, note that $\mathbf{u} \in \text{Range } A$, so there does exist a solution to $A\mathbf{v} = \mathbf{u}$ and as shown above, this solves, $A^T A\mathbf{v} = A^T\mathbf{w}$. It remains to show that any other solution $\mathbf{v}'$ to $A^T A\mathbf{v}' = A^T\mathbf{w}$ is also a least squares solution. We have the following,

$$\begin{aligned} A^T A\mathbf{v}' &= A^T\mathbf{w} \\ &= A^T A\mathbf{v} \end{aligned}$$

hence, $A^T A(\mathbf{v}' - \mathbf{v}) = \mathbf{0}$. Then by Lemma 14.18, $A(\mathbf{v}' - \mathbf{v}) = \mathbf{0}$ and hence $A\mathbf{v}' = A\mathbf{v} = \mathbf{u}$. So $\mathbf{v}'$ must also be a least squares solution. $\square$

10 Operators on the Inner Product Space

If we have a linear transformation from $V \to W$, then the adjoint is the linear transformation from $W \to V$. This will be fundemental to the special operators.

10.1 Adjoint

Definition 10.1. (*Adjoint*) Let $T : V \to W$ be a linear transformation between two finite dimensional inner product space V and W with Inner products $\langle \cdot , \cdot \rangle_V$ and $\langle \cdot , \cdot \rangle_W$ respectively. The linear operator T^* which is defined by,

$$\langle T^*(\mathbf{w}) , \mathbf{v} \rangle_V = \langle \mathbf{w} , T(\mathbf{v}) \rangle_W \qquad \forall\, \mathbf{v} \in V \text{ and } \mathbf{w} \in W$$

is called the adjoint of T.

We can also write,

$$\begin{aligned} \langle T(\mathbf{v}) , \mathbf{w} \rangle_W &= \overline{\langle \mathbf{w} , T(\mathbf{v}) \rangle_W} \\ &= \overline{\langle T^*(\mathbf{w}) , \mathbf{v} \rangle_V} \\ &= \langle \mathbf{v} , T^*(\mathbf{w}) \rangle \end{aligned}$$

this shows us that we can shift the T around and remove the star.

$$\begin{aligned} \langle T^*(\mathbf{w}) , \mathbf{v} \rangle_V &= \langle \mathbf{w} , T(\mathbf{v}) \rangle_W \\ \langle T(\mathbf{v}) , \mathbf{w} \rangle_W &= \langle \mathbf{v} , T^*(\mathbf{w}) \rangle \end{aligned}$$

We first understand the adjoint of $\mathbb{R}$ and $\mathbb{C}$ matrices, with respect to the standard inner products in $\mathbb{R}^n$ and $\mathbb{C}^n$. Hence, let $A \in A_{m\times n}(\mathbb{R})$, $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$. Then,

$$\langle \mathbf{w} , A\mathbf{v} \rangle = \begin{pmatrix} w_1 & \dots & w_m \end{pmatrix} (A) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \left((A)^T \begin{pmatrix} w_1 & \dots & w_m \end{pmatrix} \right)^T \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

or,

$$\langle \mathbf{w} , A\mathbf{v} \rangle = \mathbf{w}^T (A\mathbf{v}) = (\mathbf{w}^T A)\mathbf{v} = (A^T \mathbf{w})^T \mathbf{v} = \left\langle A^T \mathbf{w} , \mathbf{v} \right\rangle$$

Hence for real matrices, $A^* = A^T$

Now let $A \in A_{m\times n}(\mathbb{C})$, $\mathbf{v} \in \mathbb{C}^n$ and $\mathbf{w} \in \mathbb{C}^m$. Then,

$$\begin{aligned} \langle \mathbf{w} , A\mathbf{v} \rangle &= \mathbf{w}^T (\overline{A\mathbf{v}}) \\ &= (\mathbf{w}^T \overline{A}) \overline{\mathbf{v}} \\ &= (\overline{A}^T \mathbf{w})^T \overline{\mathbf{v}} \\ &= \left\langle \overline{A}^T \mathbf{w} , \mathbf{v} \right\rangle \end{aligned}$$

and hence $A^* = \overline{A}^T$. This is called the conjugate transpose.

Theorem 10.2. Let $T : V \to W$ be a linear transformation between finite dimensional innner product spaces V and W, with P an orthonormal basis of V and Q an orthonormal basis of W. Then the matrix of T^* in the Q and P bases is the conjugate transpose of the matrix of T in the P and Q bases, i.e.

$$[T^*]_Q^P = ([T]_P^Q)^* = \overline{([T]_P^Q)}^T$$

Proof. Let $A = [T]_P^Q$ and $B = [T^*]_Q^P$ and label, $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $Q = [\mathbf{w}_1, \ldots, \mathbf{w}_n]$. Then from Thm 13.18,

$$B_{ij} = \langle T^*(\mathbf{w}_j)\,,\mathbf{v}_i\rangle_V$$

$$\begin{aligned}
B_{ij} &= \langle T^*(\mathbf{w}_j)\,,\mathbf{v}_i\rangle_V \\
&= \overline{\langle \mathbf{v}_i\,, T^*(\mathbf{w}_j)\rangle}_V \\
&= \overline{\langle T(\mathbf{v}_i)\,, \mathbf{w}_j\rangle}_W \\
&= \overline{A_{ji}} \\
&= \overline{A_{ij}}^T \\
&= A^*_{ij}
\end{aligned}$$

□

Theorem 10.3. (*Properties of Adjoints*) Let $S : V \to W$ and $T : V \to W$ be linear transformations on finite dimensional inner product spaces V and W and let $c \in \mathbb{F}$. Then,

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

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

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

(iv) $(T^*)^* = T$

Theorem 10.4. (*Fundemental Theorem of Linear Algebra*) Let $T : V \to W$ be a linear transformation between finite dimensional inner product spaces V and W. Then,

(i) $\operatorname{Ker} T^* = (\operatorname{Range} T)^\perp$ and $\operatorname{Range} T = (\operatorname{Ker} T^*)^\perp$

(ii) $\operatorname{Range} T^* = (\operatorname{Ker} T)^\perp$ and $\operatorname{Ker} T = (\operatorname{Range} T^*)^\perp$

Corollary 10.5. Let $T : V \to W$ be a linear transformation between finite dimensional inner product spaces V and W. Then $\operatorname{rank} T^* = \operatorname{rank} T$

10.2 Normal and Self-Adjoint Operators

We are mostly going to consider normal operators as they are pretty neat. We see that it would be really nice to have a basis of orthonormal eigenvectors, so the operator is diagonisable and also having nice inner products. Then we have some important subclasses of the normal operators,

- self-adjoin Operators
- positive (semi-) definite operators
- unitary and orthogonal operators.

Imagine we had a normal operator, $T : V \to V$. Assume P is a basis of orthonormal eigenvectors.

$$[T]_P^P \leftarrow \text{ diagonal}$$
$$[T*]_P^P = ([T]_P^P)^* = (\overline{[T]_P^P})^T \leftarrow \text{ diagonal}$$

Now consider $[T]_P^P[T^*]_P^P = [T*]_P^P[T]_P^P$ as diagonal matrices commute. Hence,

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

This is hence the proof that,

Lemma 10.6. If V possesses an orthonormal basis of eigenvectors of T, then $TT^* = T^*T$

Definition 10.7. Let V be an Inner product space and let T be a linear operator on V. Then T is said to be normal if $TT^* = T^*T$

Now we consider the converse of above. If V is real or complex this is slightly different.

Example. Consider $T_A : \mathbb{R}^2 \to \mathbb{R}^2$

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

with $\theta \in [0, \pi]$. Now get A^*,

$$A^* = A^T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and consider $AA^* = I = A^*A$

Theorem 10.8. Let T be a linear operator on a finite dimensional complex inner product space V. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.

10.3 Self Adjoint Operators

Definition 10.9. (*Self Adjoint / Hermitian*) Let V be an inner product space and let T be a linear operator on V. Then T is said to be self-adjoint (or Hermitian) if $T = T^*$.

Remark. Self-adjoint operators are normal, since $TT^* = TT = T^*T$. If A is a self-adjoint real matrix, then $A = A^* = A^T$, i.e. A is a real symmetric matrix.

We saw above that there are some problems with normal operators and whether their eigenvalues are real or not. This isn't the case with self adjoint operators,

Lemma 10.10. Let V be a self-adjoint operator on a finite dimensional inner product space V. Then every eigenvalue of T is real.

Theorem 10.11. Let T be a linear operator on a finite dimensional inner product space V. Then T is self adjoint if and only if if there exists an orthonormal basis for V consisting of eigenvectors of T.

Adding these two together we get that; real self adjoint operators have both real eigenvalues and (orthonormal) real eigenvectors.

10.4 Positive (Semi-) Definite Operators

Definition 10.12. (*Positive (Semi-) Definite Operators*) A self-adjoint operator T on a finite dimensional inner product space V is called positive (semi-) definite if T is self-adjoint and $\langle T(\mathbf{v}), \mathbf{v}\rangle > 0$ for all $\mathbf{v} \neq \mathbf{0}_v$.

Lemma 10.13. A self-adjoint linear operator on a finite dimensional inner product space is positive (semi-) definite if and only if all of it's eigenvectors are positive (non-negative).

Given any linear operator, the operators TT^* and T^*T turn out to be positive semi-definite and are important in applications. They show up in least squares and single value decomposition.

Lemma 10.14. Let $T : V \to W$ be a linear transformation between finite dimensional inner product spaces V and W. Then,

- TT^* and T^*T are positive semi-definite operators on V and W, respectively.
- $T * T(\mathbf{v}) = 0$ if and only if $T(\mathbf{v}) = 0$.
- $\operatorname{rank}(T^*T) = \operatorname{rank}(TT^*) = \operatorname{rank} T$

Example.
- Let f be a smooth, real-valued function on $\mathbb{R}^n$. The hessian matrix $H_{ij} = \frac{\partial f}{\partial x_i \partial x_j}$ is real symmetric.
- Real symmetric matrices appear as : covariance matrices in stats, adjacency matrcies in graph theory and quadratic forms in mechanics and control theory.
- Quantum Mechanics: every physically measurable quantity is described by a self-adjoint operator, with the possible values being the eigenvalues and the corresponding quantum states being the eigenvectors.
- The positive semi-definite operators T^*T and TT^* are key to the Single Value Theorem 15.1. Also if A is real matrix then $A^*A = A^TA$ is a real symmetric matrix

10.5 Unitary and Orthogonal Operators

We now look at operators that preserve the inner product, i.e.

$$\langle T(\mathbf{u}) \, , T(\mathbf{v}) \rangle = \langle \mathbf{u} \, , \mathbf{v} \rangle$$

Assume, $\|T(\mathbf{v})\| = \|\mathbf{v}\|$, then,

$$\begin{aligned}
\langle T(\mathbf{u}) \, , T(\mathbf{v}) \rangle &= \frac{1}{2} \left(\|T(\mathbf{u}) + T(\mathbf{v})\|^2 - \|T(\mathbf{u})\|^2 - \|T(\mathbf{v})\| \right) \\
&= \frac{1}{2} \left(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\| \right)
\end{aligned}$$

Exercise. Do other direction

Now, let $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be an orthonormal basis. Then, $\langle T(\mathbf{v}_i) \, , T(\mathbf{v}_j) \rangle = \langle \mathbf{v}_i \, , \mathbf{v}_j \rangle - \delta_{ij}$. So $Q = [T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)]$ is also an orthonormal basis.

Now consider T^*T,

$$\begin{aligned}
T^*T(\mathbf{v}_j) &= \sum_{i=1}^{n} \langle T^*T(\mathbf{v}_j) \, , \mathbf{v}_i \rangle \, \mathbf{v}_i \\
&= \sum_{i=1}^{n} \langle T(\mathbf{v}_j) \, , T(\mathbf{v}_i) \rangle \, \mathbf{v}_i \\
&= \sum_{i=1}^{n} \delta_{ij} \mathbf{v}_i \\
&= \mathbf{v}_j
\end{aligned}$$

Hence, $T^*T = \mathrm{id}_V$. Since V is finite dimensional, hence $T^* = T^{-1}$.

Now let us formally define,

Definition 10.15. (*Unitary / Orthogonal Operators*) Let T be a linear operator on a finite-dimensional inner product space over the field $\mathbb{F}$. If $TT^* = T^*T = \mathrm{id}_V$, then T is called a unitary operator if $\mathbb{F} = \mathbb{C}$ and an orthognal operator if $\mathbb{F} = \mathbb{R}$.

Theorem 10.16. Let T be a linear operator on a finite-dimensional inner product space V. Then the following are equivilent,

(i) $T^*T = TT^* = \mathrm{id}_V$

(ii) $T^* = T^{-1}$

(iii) $\langle T(\mathbf{u}), T(\mathbf{v})\rangle = \langle \mathbf{u}, \mathbf{v}\rangle \,\forall\, \mathbf{u}, \mathbf{v} \in V$

(iv) If $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is an orthonormal basis for V, then $[T(\mathbf{v}_1), \ldots T(\mathbf{v}_n)]$ is an orthonormal basis for V

(v) $\|T(\mathbf{v})\| = \|\mathbf{v}\|, \forall\, \mathbf{v} \in V$.

Let us consider A^*A,

$$(A^*A)_{ij} = \overline{\mathbf{a}}_{\mathbf{i}}{}^T \mathbf{a}_j = (\overline{\mathbf{a}}_{\mathbf{i}} \mathbf{a}_j^T) = \langle \mathbf{a}_j, \mathbf{a}_i\rangle = \delta_{ij}$$

Hence $A^*A = I$.

Corollary 10.17. Let T be a linear operator on a finite dimensional inner product space V. Let P be an orthogonal basis for V and let $A = [T]_P^P$. Then A and A^* both have orthonormal columns if and only if T is a unitary or orthogonal, repsectively, operator.

Theorem 10.18. Let T be a linear operator on a finite-dimensional complex inner product space V. Then T is unitary if and only if there exists a basis V consisting of orthonormal eigenvectors of T and the eigenvalues λ_i of T all have $|\lambda_i| = 1$.

Proof. We know that if $\mathbf{v}$ is an eigenvector of $\mathbf{v}$. Then, $R(\mathbf{v}) = \lambda\mathbf{v}$. So,

$$\begin{aligned} \|T(\mathbf{v})\| &= \|\lambda\mathbf{v}\| \\ &= |\lambda|\,\|\mathbf{v}\| \end{aligned}$$

hence as, $\|T(\mathbf{v}) = \mathbf{v}\|$, then we have $|\lambda| = 1$. □

Note that if T is orthogonal, then,

$$\lambda = e^{\pm i\theta}$$

11 Single Value Decomposition

This is sometimes hailed as the greatest result in all of Linear Algebra. It improves upon diagonisation. However diagonisation is restricted to,

(i) It is restricted to square matrices

(ii) We don't always have enough eigenvectors to diagonalisation

(iii) Eigenvectors arent usually Orthogonal

(iv) There may also be problems with numerical stability

SVD solves all of these!

The single value decomposition is the factorisation of a complex/real $m \times n$ matrix A as,

$$A = U\Sigma V^*$$

where,

- U is a $m \times m$ unitary / orthogonal matrix
- V is an $n \times n$ unitary / orthogonal matrix.
- Σ is an $m \times n$ real diagonal matrix with the first r entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ positive and $\sigma_{r+1} = \sigma_k = 0$
- r is the rank of A and $k = \min\{m, n\}$

Note that is not simply a coordinate change.

$$\begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & \\ & & & \sigma_{r+1} & & & \\ & & & & \ddots & & \\ & & & & & \sigma_k & \dots \\ & & & & & \vdots & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{pmatrix}^*$$

We have names for the certain parts of the matrices,

- The $\mathbf{u}_i$'s are left singular values
- The $\mathbf{v}_i$'s are right singular values
- The σ_i's are the singular values

Since V and U are unitary/orthogonal, it is invertible with $V^{-1} = V$, then

$$A = U\Sigma V^* \iff AV = U\Sigma$$

Denoting the columns of U and V by $\mathbf{u}_i$ and $\mathbf{v}_i$ respectively, we have,

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

which reminds us of eigenvalues, but here we have different vectors on each side of the equation. This can be seen as a unit length vector $\mathbf{v}_i$ being mapped to $\mathbf{u}_i$ and being scaled by σ_i.

More generally we map a unit sphere in $\mathbb{R}^n$ to the ellipsoid in $\mathbb{R}^m$. As the basis vectors are orthonormal we stretch in orthogonal directions.

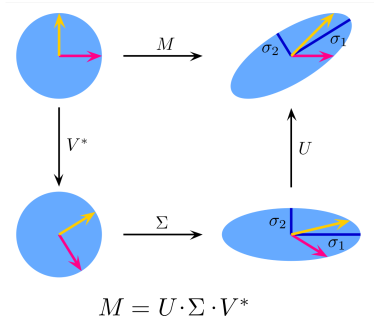

Figure 6: A diagram for Single Value Decomposition

11.1 How to find the SVD

Let us firstly look at A^*A, this matrix is semi-definite. Hence diagonalisable as it is self adjoint. It must have a basis of eigenvectors and all it's eigenvalues are non-negative.

$$
\begin{aligned}
A^*A &= (U\Sigma V^*)U\Sigma V^* \\
&= V^{**}\Sigma^* U^* U\Sigma V^* \\
&= V\Sigma^T U^{-1} U\Sigma V^* \\
&= V\Sigma^T \Sigma V^{-1}
\end{aligned}
$$

We can see that we have a square $n \times n$ matrix for $\Sigma^T\Sigma$. Then the whole matrix is diagonal, with the diagnonal of σ_i up to r.

$$
\begin{pmatrix}
\sigma_1^2 & & & & & & \\
& \sigma_2^2 & & & & & \\
& & \sigma_2^2 & & & & \\
& & & \ddots & & & \\
& & & & \sigma_k^2 & & \\
& & & & & \ddots & \\
& & & & & & 0
\end{pmatrix}
$$

We can do a similar thing for AA^*, but we will get a $m \times m$ matrix.

(i) The non-zero singular values $\sigma_1, \ldots, \sigma_r$ are obtained as the square roots of the non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$ of A^*A or AA^*. That is, $\sigma_i = \sqrt{\lambda}_i$.

(ii) The columns $\mathbf{v}_i$ of V are an orthogonal basis of $\mathbb{F}^m$ consisting of eigenvectors of A^*A.

(iii) The columns of $\mathbf{u}_i$ of U are given by $\mathbf{u}_i = \frac{1}{\sigma_i}A\mathbf{v}_i$ and are an orthonormal basis of $\mathbb{F}^m$ consisting of eigenvectors of AA^*.

The single values are unique, but the matrices U and V are not.