Math 113: Linear Algebra and Matrix Theory
Thomas Church (church@math.stanford.edu)
math.stanford.edu/~church/teaching/113/

Notes on k-wedge vectors, determinants, and characteristic polynomials

1 The space of k-wedges $\bigwedge^k V$

Definition 1.1 (Space of k-wedges)**.** Let V be a vector space, and let $k \geq 0$. Then the vector space of k-wedges from V is denoted $\bigwedge^k V$. It has the following properties:

- for every list $w_1, \ldots, w_k \in V$, there is an element $w_1 \wedge \cdots \wedge w_k \in \bigwedge^k V$ called a "k-wedge";
- the vector space $\bigwedge^k V$ is spanned by all the k-wedges $w_1 \wedge \cdots \wedge w_k$;
- the k-wedges satisfy the following three types of relations:

R1: The first relation says something like "if the symbol $\wedge$ were actually an operation, it would be distributive":

$$w_1 \wedge \cdots \wedge (aw_i + bw_i') \wedge \cdots \wedge w_k = a \cdot w_1 \wedge \cdots \wedge w_i \wedge \cdots \wedge w_k + b \cdot w_1 \wedge \cdots \wedge w_i' \wedge \cdots \wedge w_k$$

For example, this is what tells us that we can "FOIL" out a product

$$\begin{aligned}(u+v) \wedge (w+x) &= u \wedge (w+x) + v \wedge (w+x) && \text{applying R1 to the first term}\\ &= u \wedge w + u \wedge x + v \wedge w + v \wedge x && \text{applying R1 to the second term}\end{aligned}$$

This is also what tells us that if one of your k vectors is the zero vector $\vec{0} \in V$, then the k-wedge is zero in $\bigwedge^k V$:

$$\begin{aligned}w_1 \wedge \cdots \wedge \vec{0} \wedge \cdots \wedge w_k &= w_1 \wedge \cdots \wedge (0 \cdot \vec{0}) \wedge \cdots \wedge w_k\\ &= 0 \cdot (w_1 \wedge \cdots \wedge \vec{0} \wedge \cdots \wedge w_k) = \vec{0} \in \textstyle\bigwedge^k V\end{aligned}$$

R2: If two k-wedges involve the exact same list of k vectors, but in a different order (like u, v, w versus w, v, u), the second relation says the first k-wedge is ± 1 times the second k-wedge. For example:

$$u \wedge v \wedge w = -v \wedge u \wedge w$$

Whether the sign is $+1$ or -1 depends on how the vectors are rearranged. The simplest way to describe this is to say that if just *two* vectors in the list are swapped, then the k-wedges are negatives of each other:

$$w_1 \wedge \cdots \wedge w_i \wedge \cdots \wedge w_j \wedge \cdots \wedge w_k = -w_1 \wedge \cdots \wedge w_j \wedge \cdots \wedge w_i \wedge \cdots \wedge w_k$$

If you have a more complicated rearrangement, you can figure out whether the sign is $+$ or $-$ by applying this rule multiple times. For example, if we want to compare $u \wedge v \wedge w$ with $v \wedge w \wedge u$:

$$u \wedge v \wedge w = -\, v \wedge u \wedge w \qquad \text{(applying R2 to switch first two)}$$
$$v \wedge u \wedge w = -v \wedge w \wedge u \qquad \text{(applying R2 to switch last two)}$$

so combining these we have

$$u \wedge v \wedge w = -v \wedge u \wedge w = v \wedge w \wedge u$$

R3: If your list of k vectors has a repetition, then the resulting k-wedge is equal to zero:

$$w_1 \wedge \cdots \wedge w_k = 0 \qquad \text{if two entries } w_i \text{ and } w_j \text{ are equal}$$

For example, $u \wedge v \wedge u$ is always $= 0$, since the vector u occurs twice in the list.

- finally, there are no "hidden relations"; any relation that holds in $\bigwedge^k V$ follows from those relations listed above.

2 Wedge Dependence and Independence

There are two key theorems about $\bigwedge^k V$ that we will find very useful.

Theorem 2.1 (**Wedge Dependence Lemma).** Let $w_1, \ldots, w_k$ be any list of k vectors in V. Then

$$w_1 \wedge \cdots \wedge w_k = 0 \in \bigwedge^k V \qquad \iff \qquad w_1, \ldots, w_k \text{ is linearly dependent in } V$$

Theorem 2.2 (**Wedge Independence Lemma).** Let $\mathcal{B} = (v_1, \ldots, v_n)$ be a basis for V. Then a basis for $\bigwedge^k V$ is given by the set of "basic k-wedges"

$$\mathcal{B}_k = \left\{\ v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_k} \ \middle|\ 1 \le i_1 < i_2 < \cdots < i_k \le n\ \right\}.$$

Proof of the Wedge Dependence Lemma (using WIL).

($\Longleftarrow$) One direction of the Wedge Dependence Lemma follows from the Wedge Independence Lemma, as follows. If $w_1, \dots, w_k$ are linearly *independent*, then we can extend this list to a basis $\mathcal{B} = (w_1, \dots, w_n)$ of V. The Wedge Independence Lemma tells us the set $\mathcal{B}_k$ is a basis for $\bigwedge^k V$. But the very first element of $\mathcal{B}_k$ is the basic k-wedge $w_1 \wedge w_2 \wedge \cdots \wedge w_k$. Since no element of a basis can ever be $= 0$, this implies that $w_1 \wedge \cdots \wedge w_k \neq 0$.

($\Longrightarrow$) Now assume that $w_1, \dots, w_k$ are linearly dependent. To simplify the proof, assume that w_k is linearly dependent on the other vectors. (If this is not true, we can make it true by switching w_k with some vector w_ℓ that *is* linearly dependent on the others; by R2 this multiplies the k-wedge by -1, which will not affect whether or not it is $= 0$ or not.)

Therefore we can write $w_k = c_1 w_1 + c_2 w_2 + \cdots + c_{k-1} w_{k-1}$. Therefore distributing using R1 we have

$$\begin{aligned} w_1 \wedge \cdots \wedge w_{k-1} \wedge w_k &= w_1 \wedge \cdots \wedge w_{k-1} \wedge (c_1 w_1 + c_2 w_2 + \cdots + c_{k-1} w_{k-1}) \\ &= \sum_{i=1}^{k-1} w_1 \wedge \cdots \wedge w_{k-1} \wedge c_i w_i \\ &= \sum_{i=1}^{k-1} c_i w_1 \wedge \cdots \wedge w_{k-1} \wedge w_i \end{aligned}$$

In each term $w_1 \wedge \cdots \wedge w_{k-1} \wedge w_i$ the vector w_i is repeated, so by R3 this k-wedge is zero. Since every term in the sum is zero, $w_1 \wedge \cdots \wedge w_{k-1} \wedge w_k = 0$, as desired. □

Proof of the Wedge Independence Lemma. We first need to prove that $\mathcal{B}_k$ spans $\bigwedge^k V$. The formal proof is below, but first let's work out a concrete example in full detail:

Let's say that V is 3-dimensional, so that $\mathcal{B} = (v_1, v_2, v_3)$ is a basis for V. Then for $k = 2$, the set $\mathcal{B}_2$ is just

$$\begin{aligned} \mathcal{B}_2 &= \{ \ v_i \wedge v_j \mid 1 \leq i < j \leq 3 \ \}. \\ &= \{ \ v_1 \wedge v_2, \quad v_1 \wedge v_3, \quad v_2 \wedge v_3 \ \} \end{aligned}$$

Let w and u be any vectors in V; we want to prove that $w \wedge u \in \operatorname{span}(\mathcal{B}_2)$. Since v_1, v_2, v_3 is a basis, we can write $w = a_1 v_1 + a_2 v_2 + a_3 v_3$ and $u = b_1 v_1 + b_2 v_2 + b_3 v_3$. Therefore by applying R1 twice, we have:

$$
\begin{aligned}
w \wedge u = & (a_1 v_1 + a_2 v_2 + a_3 v_3) \wedge (b_1 v_1 + b_2 v_2 + b_3 v_3) \\
= & \; a_1 \cdot v_1 \wedge (b_1 v_1 + b_2 v_2 + b_3 v_3) \\
& + a_2 \cdot v_2 \wedge (b_1 v_1 + b_2 v_2 + b_3 v_3) \\
& + a_3 \cdot v_3 \wedge (b_1 v_1 + b_2 v_2 + b_3 v_3) \\
= & \; a_1 b_1 \cdot v_1 \wedge v_1 \; + \; a_1 b_2 \cdot v_1 \wedge v_2 \; + \; a_1 b_3 \cdot v_1 \wedge v_3 \\
& + a_2 b_1 \cdot v_2 \wedge v_1 \; + \; a_2 b_2 \cdot v_2 \wedge v_2 \; + \; a_2 b_3 \cdot v_2 \wedge v_3 \\
& + a_3 b_1 \cdot v_3 \wedge v_1 \; + \; a_3 b_2 \cdot v_3 \wedge v_2 \; + \; a_3 b_3 \cdot v_3 \wedge v_3
\end{aligned}
$$

Three of the terms here are already multiples of elements of $\mathcal{B}_2$:

$$
\begin{array}{ccccc}
a_1 b_2 \cdot v_1 \wedge v_2 & a_1 b_3 \cdot v_1 \wedge v_3 & & \text{OK} & \text{OK} \\
 & a_2 b_3 \cdot v_2 \wedge v_3 & & & \text{OK}
\end{array}
$$

But what can we do about the other six terms? Well, the relation R3 tells us that $v_1 \wedge v_1$ is just equal to 0, since it has a repeated vector. Similarly $v_2 \wedge v_2 = 0$ and $v_3 \wedge v_3 = 0$. Therefore we can just drop those three terms from the sum, since they'll be equal to zero.

$$
\begin{array}{ccccccc}
a_1 b_1 \cdot v_1 \wedge v_1 & & & & 0 & & \\
& + \; a_2 b_2 \cdot v_2 \wedge v_2 & & = & & +0 & \\
& & + \; a_3 b_3 \cdot v_3 \wedge v_3 & & & & +0
\end{array}
$$

For the last three terms, the relation R2 tells us that e.g. $v_3 \wedge v_1$ is equal to $-v_3 \wedge v_1$, since swapping two vectors negates a k-wedge. Therefore we can replace them:

$$
\begin{array}{lcl}
a_2 b_1 \cdot v_2 \wedge v_1 & = & -\, a_2 b_1 \cdot v_1 \wedge v_2 \\
a_3 b_1 \cdot v_3 \wedge v_1 \; + \; a_3 b_2 \cdot v_3 \wedge v_2 & & -\, a_3 b_1 \cdot v_1 \wedge v_3 - a_3 b_2 \cdot v_2 \wedge v_3
\end{array}
$$

Therefore we can simplify the sum above as

$$\begin{aligned} w \wedge u = \ & a_1b_1 \cdot v_1 \wedge v_1 \ + \ a_1b_2 \cdot v_1 \wedge v_2 \ + \ a_1b_3 \cdot v_1 \wedge v_3 \\ & + a_2b_1 \cdot v_2 \wedge v_1 \ + \ a_2b_2 \cdot v_2 \wedge v_2 \ + \ a_2b_3 \cdot v_2 \wedge v_3 \\ & + a_3b_1 \cdot v_3 \wedge v_1 \ + \ a_3b_2 \cdot v_3 \wedge v_2 \ + \ a_3b_3 \cdot v_3 \wedge v_3 \\ \\ = \ & (a_1b_2 - b_2a_1) \cdot v_1 \wedge v_2 \ + \ (a_1b_3 - b_3a_1) \cdot v_1 \wedge v_3 \\ & \qquad + \ (a_2b_3 - a_3b_2) \cdot v_2 \wedge v_3 \end{aligned}$$

This shows that $w \wedge u \in \operatorname{span}(v_1 \wedge v_2, \quad v_1 \wedge v_3, \quad v_2 \wedge v_3)$, as desired.

The argument in the general case is the same. We know that $\bigwedge^k V$ is spanned by k-wedges $w_1 \wedge \cdots \wedge w_k$, so it suffices to prove that any such k-wedge is in the span of $\mathcal{B}_k$. Since $v_1, \ldots, v_n$ is a basis for V, we can write each w_i as a linear combination of the basis vectors $v_1, \ldots, v_n$. If we then expand out $w_1 \wedge \cdots \wedge w_k$ using the distributivity from R1, each term in the resulting sum will be some multiple of a k-wedge of the form

$$v_{i_1} \wedge \cdots \wedge v_{i_k} \qquad \text{for } 1 \le i_1 \le n, \quad \ldots, \quad 1 \le i_k \le n. \tag{1}$$

This is already enough to show that $\bigwedge^k V$ is spanned by the n^k fairly-simple k-wedges of the form (1). So we just need to show that each of these is in the span of $\mathcal{B}_k$.

Many of the k-wedges (1) are actually zero: if there is any repeated index (meaning $i_j = i_\ell$ for some $j \neq \ell$), the k-wedge is zero by R2. Therefore it remains only to handle those k-wedges (1) where the indices $i_1, \ldots, i_k$ are all distinct. If the indices are *not* in order, then we could rearrange the vectors in the k-wedge (1) so that they are in order; by R2 this multiplies the k-wedge by ± 1, so it does not change whether or not the k-wedge lies in $\operatorname{span}(\mathcal{B}_k)$. But this reduces us to those k-wedges (1) where the indices are in order (meaning $i_1 < i_2 < \cdots < i_k$)— and this is exactly our purported generating set $\mathcal{B}_k$! This completes the proof that $\bigwedge^k V$ is spanned by $\mathcal{B}_k$. □

The other half of the Wedge Independence Lemma is to prove that $\mathcal{B}_k$ is linearly independent. However, this argument is quite complicated and somewhat unenlightening, so we won't do it in Math 113. The idea is to show that no linear dependence between the vectors in $\mathcal{B}_k$ can possibly follow from the relations R1, R2, and R3 (since we know that all relations in $\bigwedge^k V$ follow from these). But there are infinitely many relations of the form R1, R2, and R3, so it requires a fair bit of work to show that *no* linear dependence can follow from them. I'm happy to talk about this with anyone who's interested, but for class and the homeworks, please just take my word for it.

The dimension of $\bigwedge^k V$. Since the Wedge Independence Lemma gives us a basis for $\bigwedge^k V$, we can find the dimension of $\bigwedge^k V$.

Corollary 2.3. If $\dim V = n$, then the dimension of $\bigwedge^k V$ is $\binom{n}{k}$, the number of ways of choosing a k-element subset from $\{1, \ldots, n\}$.

Proof. The dimension $\dim \bigwedge^k V$ is the number of elements in any basis. If $\mathcal{B} = (v_1, \ldots, v_n)$, the Wedge Independence Lemma tells us that

$$\mathcal{B}_k = \left\{ \, v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_k} \;\middle|\; 1 \leq i_1 < i_2 < \cdots < i_k \leq n \, \right\}$$

is a basis for $\bigwedge^k V$, so we just need to count how many k-wedges are in $\mathcal{B}_k$. The number of k-wedges in this basis is the number of ways to pick k numbers $i_1, i_2, \ldots, i_k$ between 1 and n that are in increasing order. But if we are given k different numbers between 1 and n, there is a unique way to *put* them in increasing order. Therefore we get one basis element for each subset $\{i_1, i_2, \ldots, i_k\} \subset \{1, \ldots, n\}$ of size k— and the number of such subsets is $\binom{n}{k}$ by definition. □

Since $\binom{n}{n} = 1$ (there's only one way to pick an n-element subset of $\{1, \ldots, n\}$), this gives the following fact.

Corollary 2.4. If $\dim V = n$, then $\bigwedge^n V$ is 1-dimensional.

Moreover if k is **greater than** n, there is *no* way to pick numbers $1 \leq i_1 < \cdots < i_k \leq n$. Therefore $\mathcal{B}_k$ is the empty set, and so $\bigwedge^k V$ must be the zero vector space.

Corollary 2.5. If $\dim V = n$ and $k > n$, then $\bigwedge^k V = \{0\}$.

3 Determinants

Definition 3.1 (Induced operator T_*)**.** If $T \colon W \to U$ is a linear transformation, there is an "induced transformation" $T_* \colon \bigwedge^k W \to \bigwedge^k U$ defined[1] on k-wedges by

$$T_*(w_1 \wedge \cdots \wedge w_k) = T(w_1) \wedge \cdots \wedge T(w_k)$$

In particular, if $T \in \mathcal{L}(V)$ is an operator on V, there is an induced operator $T_* \colon \bigwedge^k V \to \bigwedge^k V$ defined by the above formula.

[1] One slightly tricky point is showing that this formula is well-defined, i.e. gives an unambiguous linear transformation $T_* \colon \bigwedge^k W \to \bigwedge^k U$. This just requires checking that T_* preserves the relations R1, R2, and R3, which follows from the linearity of T. However I will skip the details of this computation.

Determinant. If $\dim V = n$, for any operator $T \in \mathcal{L}(V)$ we have the induced operator $T_*\colon \bigwedge^n V \to \bigwedge^n V$. By Corollary 2.4, $\bigwedge^n V$ is 1-dimensional, so if $\omega \in \bigwedge^n V$ is any nonzero element, we must have $T_*(\omega) = d \cdot \omega$ for some unique $d \in \mathbb{F}$. This leads to the general mathematical definition of the determinant:

Definition 3.2. If $\dim V = n$ and $T \in \mathcal{L}(V)$, the determinant $\det(T) \in \mathbb{F}$ is the unique number such that

$$T_*(\omega) = \det(T) \cdot \omega \qquad \text{for any } \omega \in \textstyle\bigwedge^n V.$$

Concretely, if $v_1, \ldots, v_n$ is any basis for V, then we have

$$T(v_1) \wedge \cdots \wedge T(v_n) = \det(T) \cdot v_1 \wedge \cdots \wedge v_n. \tag{$*$}$$

We can now verify all the standard properties of the determinant.

The determinant does not depend on basis. The determinant of T does not depend on the basis $v_1, \ldots, v_n$ we choose; this is implicit in the definition above, but we can check it directly. If we picked a different basis $w_1, \ldots, w_n$, since $\dim \bigwedge^n V = 1$ we know that $w_1 \wedge \cdots \wedge w_n = c \cdot v_1 \wedge v_n$ for some $c \in \mathbb{F}$. Therefore

$$\begin{aligned} T_*(w_1 \wedge \cdots \wedge w_n) &= T_*(c \cdot v_1 \wedge \cdots \wedge v_n) \\ &= c \cdot T_*(v_1 \wedge \cdots \wedge v_n) \\ &= c \cdot \det(T) \cdot v_1 \wedge \cdots \wedge v_n \\ &= \det(T) \cdot (c \cdot v_1 \wedge \cdots \wedge v_n) \\ &= \det(T) \cdot w_1 \wedge \cdots \wedge w_n \end{aligned}$$

This shows that we can compute the determinant using *any* basis for V that we like.

Invertibility and determinants.

Proposition 3.3. If $\dim V = n$ and $T \in \mathcal{L}(V)$, then

$$\det(T) \neq 0 \quad \iff \quad T \text{ is invertible.}$$

Proof. Let $v_1, \ldots, v_n$ be a basis for V. Since $v_1, \ldots, v_n$ is linearly independent, the Wedge Dependence Lemma tells us that $v_1 \wedge \cdots \wedge v_n \neq 0 \in \bigwedge^n V$. We know that $T(v_1) \wedge \cdots \wedge T(v_n) = \det(T) \cdot v_1 \wedge \cdots \wedge v_n$, so

$$\det(T) \neq 0 \quad \iff \quad T(v_1) \wedge \cdots \wedge T(v_n) \neq 0.$$

However, the Wedge Dependence Lemma tells us that

$$T(v_1) \wedge \cdots \wedge T(v_n) \neq 0 \quad \iff \quad T(v_1), \ldots, T(v_n) \text{ are linearly independent.}$$

Question 1 on the midterm states that

$$T(v_1), \ldots, T(v_n) \text{ are linearly independent} \quad \Longleftrightarrow \quad T \text{ is injective.}$$

Finally, since T is an operator on V, Theorem 3.21 from the book says that

$$T \text{ is injective} \quad \Longleftrightarrow \quad T \text{ is invertible.} \qquad \square$$

Explicit formula for $\det(T)$ **when** $\dim V = 2$**.** Assume $\dim V = 2$ and v_1, v_2 is a basis for V, and that $T \in \mathcal{L}(V)$ satisfies $T(v_1) = av_1 + cv_2$ and $T(v_2) = bv_1 + dv_2$. (The point of this labeling is that the matrix of T w.r.t. the basis v_1, v_2 is $M(T) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.) Then:

$$\begin{aligned}
T_*(v_1 \wedge v_2) &= T(v_1) \wedge T(v_2) \\
&= (av_1 + cv_2) \wedge (bv_1 + dv_2) \\
&= ab \cdot v_1 \wedge v_1 + ad \cdot v_1 \wedge v_2 + bc \cdot v_2 \wedge v_1 + ad \cdot v_2 \wedge v_2 \\
&= ad \cdot v_1 \wedge v_2 + bc \cdot v_2 \wedge v_1 \\
&= ad \cdot v_1 \wedge v_2 - bc \cdot v_1 \wedge v_2 \\
&= (ad - bc) v_1 \wedge v_2
\end{aligned}$$

By the defining property $(*)$ of the determinant this implies $\det(T) = ad - bc$. In particular, we find that a 2-by-2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad - bc \neq 0$.

Determinant is multiplicative.

Proposition 3.4. If $T \in \mathcal{L}(V)$ and $S \in \mathcal{L}(V)$ are operators on V, then

$$\det(ST) = \det(S) \cdot \det(T) = \det(TS).$$

Proof. We first note that $(S \circ T)_* = (S_*) \circ (T_*) \in \mathcal{L}(\bigwedge^n V)$:

$$\begin{aligned}
(S_*) \circ (T_*)(v_1 \wedge \cdots \wedge v_n) &= (S_*)(T(v_1) \wedge \cdots \wedge T(v_n)) \\
&= S(T(v_1)) \wedge \cdots \wedge S(T(v_n)) \\
&= (S \circ T)_*(v_1 \wedge \cdots \wedge v_n)
\end{aligned}$$

Therefore for any $\omega \in \bigwedge^n V$ we have

$$(S \circ T)_*(\omega) = S_*(T_*(\omega)) = S_*(\det(T) \cdot \omega) = \det(T) \cdot S_*(\omega) = \det(T) \cdot \det(S) \cdot \omega. \qquad \square$$

Determinant of a diagonalizable operator is product of eigenvalues.

Proposition 3.5. If $T \in \mathcal{L}(V)$ is diagonalizable, let $v_1, \ldots, v_n$ be a basis of eigenvectors, so that $T(v_i) = \lambda_i \cdot v_i$. Then $\det(T) = \lambda_1 \lambda_2 \cdots \lambda_n$ is the product of these eigenvalues.

Proof.

$$\begin{aligned} T(v_1) \wedge T(v_2) \wedge \cdots \wedge T(v_n) &= (\lambda_1 v_1) \wedge (\lambda_2 v_2) \wedge \cdots \wedge (\lambda_n v_n) \\ &= (\lambda_1 \lambda_2 \cdots \lambda_n) \cdot v_1 \wedge v_2 \wedge \cdots \wedge v_n. \end{aligned}$$

□

The simplicity of this proof shows how useful it is that we can compute $\det(T)$ using *any* basis of V.

Determinant of an upper-triangular operator is product of diagonal entries.

Definition 3.6. Let $v_1, \ldots, v_n$ be a basis of V. We say that an operator $T \in \mathcal{L}(V)$ is *upper-triangular* with respect to this basis

$$T(v_i) \in \operatorname{span}(v_1, \ldots, v_i) \quad \text{for all } i = 1, \ldots, n.$$

In this case we can write

$$T(v_i) = d_i \cdot v_i + w_i \qquad \text{for some } d_i \in \mathbb{F} \text{ and } w_i \in \operatorname{span}(v_1, \ldots, v_{i-1}).$$

Proposition 3.7. If $T \in \mathcal{L}(V)$ is upper-triangular with respect to the basis $v_1, \ldots, v_n$, write $T(v_i) = d_i \cdot v_i + w_i$ for $w_i \in \operatorname{span}(v_1, \ldots, v_{i-1})$. Then $\det(T) = d_1 d_2 \cdots d_n$ is the product of these "diagonal entries".

Proof.

$$\begin{aligned} T(v_1) \wedge T(v_2) \wedge \cdots \wedge T(v_n) &= (d_1 v_1) \wedge (d_2 v_2 + w_2) \wedge (d_3 v_3 + w_3) \wedge \cdots \wedge (d_n v_n + w_n) \\ &= d_1 d_2 v_1 \wedge v_2 \wedge (d_3 v_3 + w_3) \wedge \cdots \wedge (d_n v_n + w_n) \\ &\quad + d_1 v_1 \wedge w_2 \wedge (d_3 v_3 + w_3) \wedge \cdots \wedge (d_n v_n + w_n) \\ \text{however since } w_2 \in \operatorname{span}(v_1), &\text{ the vectors in the second } n\text{-wedge are linearly dependent,} \\ &\text{so this second } n\text{-wedge is 0 by the WDL} \\ &= d_1 d_2 v_1 \wedge v_2 \wedge (d_3 v_3 + w_3) \wedge \cdots \wedge (d_n v_n + w_n) \\ &= d_1 d_2 d_3 v_1 \wedge v_2 \wedge v_3 \wedge \cdots \wedge (d_n v_n + w_n) \\ &\quad + d_1 d_2 v_1 \wedge v_2 \wedge w_3 \wedge \cdots \wedge (d_n v_n + w_n) \\ \text{however since } w_w \in \operatorname{span}(v_1, v_2), &\text{ the vectors in the second } n\text{-wedge are linearly dependent,} \\ &\text{so this second } n\text{-wedge is 0 by the WDL} \\ &= \qquad \vdots \\ &= (d_1 d_2 \cdots d_n) \cdot v_1 \wedge v_2 \wedge \cdots \wedge v_n. \end{aligned}$$

□

Proposition 3.8. Each "diagonal entry" d_j of an upper-triangular operator T is an eigenvalue of T, and each eigenvalue of T appears as one of the diagonal entries.

Proof. We assume that T is upper-triangular w.r.t. the basis $v_1, \ldots, v_n$, and write$T(v_i) = d_i \cdot v_i + w_i$ as above. Given $\lambda \in \mathbb{F}$, consider the operator $T - \lambda I$. We have

$$(T - \lambda I)(v_i) = T(v_i) - \lambda \cdot v_i = (d_i - \lambda) \cdot v_i + w_i.$$

Therefore $T - \lambda I$ is also upper-triangular w.r.t. the basis $v_1, \ldots, v_n$, and the "diagonal entries" of $T - \lambda I$ are the coefficients $d_i - \lambda$ for $i = 1, \ldots, n$. By Proposition 3.7, we have

$$\det(T - \lambda I) = (d_1 - \lambda)(d_2 - \lambda) \cdots (d_n - \lambda).$$

We already know that λ is an eigenvalue of $T \iff \det(T - \lambda I) = 0$, and from this formula we see that this holds exactly when λ is equal to one of the diagonal entries d_i. (Otherwise each factor $d_i - \lambda$ would be nonzero, so the product $\det(T)$ would be nonzero as well.) $\square$

4 Characteristic polynomials

Definition 4.1. Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional vector space V. Then the *characteristic polynomial* $\chi_T(x)$ is the function (of $x \in \mathbb{F}$) defined by

$$\chi_T(x) = \det(xI - T).$$

The characteristic polynomial $\chi_T(x)$ is a polynomial in x (with coefficients in $\mathbb{F}$); if $\dim V = n$, then $\chi_T(x)$ has degree n.

Proposition 4.2. The roots of $\chi_T(x)$, meaning the set of $\lambda \in \mathbb{F}$ such that $\chi_T(\lambda) = 0$, are the eigenvalues of T.

Proof. Note that $xI - T$ is just (-1) times $T - xI$, so $\det(xI - T) = \pm\det(T - xI)$. Therefore

$$\begin{aligned}
\chi_T(\lambda) = 0 &\iff \det(xI - T) = 0 \\
&\iff \det(T - xI) = 0 \\
&\iff T - xI \text{ is not invertible, by Prop 3.3} \\
&\iff T - xI \text{ is not injective} \\
&\iff \exists v \neq 0 \in V \text{ s.t. } (T - xI)(v) = 0,
\end{aligned}$$

and this last condition says precisely that v is a nonzero eigenvector with eigenvalue λ. $\square$

Theorem 4.3 (Cayley–Hamilton Theorem)**.** $\chi_T(T) = 0$.

The Cayley–Hamilton theorem finally gives us the following corollary, which we have been waiting a long time for.

Corollary 4.4. The degree of the minimal polynomial $m_T(x)$ is $\leq \dim V$.

Proof. Since $\chi_T(T) = 0$ by the Cayley–Hamilton theorem, the degree of the minimal polynomial must be $\leq$ the degree of $\chi_T(x)$ [otherwise it wouldn't be minimal!]. Since $\deg \chi_T(x) = \dim V$, this shows that $\deg m_T(x) \leq \dim V$. $\square$

Although the Cayley–Hamilton theorem does hold over arbitrary fields (including weird ones like $\mathbb{F}_2$), we will only prove it for real and complex operators. For complex operators this will be easy, once we know that every complex operator is upper-triangular w.r.t. some basis.

Proposition 4.5. If $T \in \mathcal{L}(V)$ is an operator on a finite-dimensional vector space V over $\mathbb{C}$, there exists a basis $v_1, \ldots, v_n$ for V w.r.t. which T is upper-triangular.

Proof. We prove this proposition by induction on the dimension of V. The base case is when $\dim V = 1$, in which case $T = \lambda I$ for some λ, so any nonzero vector $v_1 \in V$ works.

Now let $\dim V = n$; by induction, we can assume that Proposition 4.5 (this proposition!) can be safely applied to any complex vector space of dimension $< n$. Our goal is to find a basis $v_1, \ldots, v_n$ so that $T(v_i) \in \operatorname{span}(v_1, \ldots, v_i)$ for all $i = 1, \ldots, n$.

Since T is a complex operator, we know that T has at least one eigenvalue. So let λ be some eigenvalue of T, and let $U = \operatorname{Image}(T - \lambda I)$. Note that for any $v \in V$ we have the decomposition (essentially tautologically)

$$T(v) = \lambda v + (T - \lambda I)v$$

where the second term $(T-\lambda I)v$ is obviously contained in $U = \operatorname{Image}(T-\lambda I)$. In particular, this shows that the subspace U is invariant under T, since for any $u \in U$ we have

$$T(u) = \lambda u + (T - \lambda I)u \in U$$

(the first term lies in U because $u \in U$, and the second term lies in U by definition).

So far this hasn't even used that λ is an eigenvalue! So let's use that. Since λ is an eigenvalue of T, we know that $\ker(T - \lambda I) \neq \{0\}$. Applying the Rank-Nullity theorem to $T - \lambda I$, this implies that $\operatorname{Image}(T - \lambda I) \neq V$. In particular, $\dim U$ is **strictly** smaller than $\dim V$. Therefore by our inductive hypothesis we may apply Proposition 4.5 to the operator $T|_U \colon U \to U$. This gives us a basis $v_1, \ldots, v_k$ of U so that $U = \operatorname{span}(v_1, \ldots, v_k)$ and

$$T|_U(v_i) \in \operatorname{span}(v_1, \ldots, v_i) \quad \text{for all } i = 1, \ldots, k.$$

Now extend this basis to a basis $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ for V. Since $T(v_i) = T|_U(v_i)$, we've already verified the condition

$$T(v_i) \in \operatorname{span}(v_1, \ldots, v_i) \qquad \text{for all } i = 1, \ldots, k.$$

And for the remaining basis vectors v_i with $i = k+1, \ldots, n$ we have the tautological equation

$$T(v_i) = \lambda v_i + (T - \lambda I)v_i.$$

Since $(T - \lambda I)v_i \in U$ by definition of U, and $U = \operatorname{span}(v_1, \ldots, v_k)$, this shows that

$$T(v_i) \in \operatorname{span}(v_1, \ldots, v_k, v_i).$$

The span $\operatorname{span}(v_1, \ldots, v_k, v_i)$ is automatically contained in the larger span $\operatorname{span}(v_1, \ldots, v_k, v_{k+1}, \ldots, v_i)$, so we have that

$$T(v_i) \in \operatorname{span}(v_1, \ldots, v_k, v_i) \subset \operatorname{span}(v_1, \ldots, v_k, v_{k+1}, \ldots, v_i).$$

Since for all $i = 1, \ldots, n$ we have $T(v_i) \in \operatorname{span}(v_1, \ldots, v_i)$, the operator T is upper-triangular w.r.t. the basis $v_1, \ldots, v_n$ that we have constructed. □

We can now prove the Cayley–Hamilton theorem for complex operators.

Proof of Cayley–Hamilton Theorem over $\mathbb{C}$. Sketch of proof: Find an upper-triangular basis with

$$T(v_i) = d_i \cdot v_i + w_i.$$

The operator $xI - T$ is still upper-triangular, with diagonal entries $x - d_1, \ldots, x - d_n$. Therefore by Proposition 3.7,

$$\chi_T(x) = \det(xI - T) = (x - d_1) \cdots (x - d_n).$$

Therefore our goal is to prove that

$$(T - d_1 I) \cdots (T - d_n I) = 0.$$

Define $U_i = \operatorname{span}(v_1, \ldots, v_i)$. Since $T - d_i I$ is upper-triangular, we know that $(T - d_i I)(v_j) \in U_j$ for any j. However, we also have that

$$(T - d_i I)(v_i) = T(v_i) - d_i \cdot v_i = (d_i \cdot v_i + w_i) - d_i \cdot v_i = w_i$$

for $w_i \in U_{i-1}$. This implies that

$$(T - d_i I)(U_i) \subset U_{i-1} \qquad \text{for all } i \tag{$*$}$$

Since $V = U_n$, taking $i = n$ in $(*)$ implies that $(T - d_nI)(V) \subset U_{n-1}$. Therefore applying $(*)$ again with $i = n-1$ shows that that $(T - d_{n-1}I)(T - d_nI)(V) \subset U_{n-2}$; and so on. Repeatedly applying $(*)$ shows that

$$(T - d_kI)(T - d_{k+1}I)\cdots(T - d_nI)(V) \subset U_{k-1}.$$

Therefore after applying $(*)$ a full n times, we have

$$(T - d_1I)(T - d_2I)\cdots(T - d_nI)(V) \subset U_0 = \{0\}.$$

This shows that $(T - d_1I)\cdots(T - d_nI) = 0$, as desired. □

5 Complexification

If V is a vector space over $\mathbb{R}$, we can define its *complexification* $V_{\mathbb{C}}$, which is a complex vector space, as follows. The elements of $V_{\mathbb{C}}$ are pairs (v, w) where $v \in V$ and $w \in W$, with addition defined component-wise: $(v, w) + (v', w') = (v + v', w + w')$. Scalar multiplication by real numbers $a \in \mathbb{R}$ is similarly easy: we set $a \cdot (v, w) = (a \cdot v, a \cdot w)$.

However, to make $V_{\mathbb{C}}$ into a vector space over $\mathbb{C}$, we need to define scalar multiplication by *complex* numbers, not just real. The key is to define multiplication by $i \in \mathbb{C}$, which we define by

$$i \cdot (v, w) = (-w, v).$$

Since every complex number can be written as $a + bi$, and scalar multiplication is supposed to be distributive, this tells us how we must define multiplication by $a + bi$:

$$(a + bi) \cdot (v, w) = (a \cdot v - b \cdot w, \quad b \cdot v + a \cdot w).$$

This is what we take as the formal definition of scalar multiplication; however multiplication by i is really the heart of it. Anyway, you can check without difficulty that this definition makes $V_{\mathbb{C}}$ into a vector space over $\mathbb{C}$.

Notation $v + iw$ for elements of $V_{\mathbb{C}}$. There is a more natural notation that we can use for the elements of $V_{\mathbb{C}}$. Let us slightly abuse notation and denote the element $(v, 0) \in V_{\mathbb{C}}$ simply as v. If we make this convention, then $i \cdot v$ would be the element

$$i \cdot v = i \cdot (v, 0) = (-0, v) = (0, v)$$

by applying the definition above. Therefore instead of writing $(v, w) \in V_{\mathbb{C}}$, we can denote this element by $v + iw$, since

$$(v, w) = (v, 0) + (0, w) = v + iw.$$

This leads us to think of $v \in V$ as the "real part" of $v + iw \in V_{\mathbb{C}}$, and $w \in V$ as the "imaginary part"; this analogy will not have many mathematical consequences for us, but it is a useful metaphor.

Proposition 5.1. If $v_1, \ldots, v_n$ is a basis for V, then $v_1, \ldots, v_n$ is a basis for $V_{\mathbb{C}}$.

Proof. Consider an arbitrary element $(v, w) = v + iw \in V_{\mathbb{C}}$. We know that v can be uniquely written as $v = a_1 v_1 + \cdots + a_n v_n$, and w can be uniquely written as $w = b_1 v_1 + \cdots + b_n v_n$. Set $z_1 = a_1 + ib_1 \in \mathbb{C}, \ldots, z_n = a_n + ib_n \in \mathbb{C}$. Then we can write $v + iw$ as

$$\begin{aligned} v + iw &= (a_1 v_1 + \cdots + a_n v_n) + i(b_1 v_1 + \cdots + b_n v_n) \\ &= (a_1 + ib_1)v_1 + \cdots + (a_n + ib_n)v_n \\ &= z_1 v_1 + \cdots + z_n v_n, \end{aligned}$$

and it is easy to check that this representation is unique. □

Complexification of linear transformations. If V and U are real vector spaces and $T\colon V \to U$ is a linear transformation, we can define a linear transformation

$$T_{\mathbb{C}}\colon V_{\mathbb{C}} \to U_{\mathbb{C}}$$

$$\text{by} \quad T_{\mathbb{C}}(v + iw) = T(v) + iT(w).$$

This will be most useful for us when $T \in \mathcal{L}(V)$ is a real operator on V, in which case $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ is a complex operator on the complexification $V_{\mathbb{C}}$. The complexified operator $T_{\mathbb{C}}$ has some elementary properties that will be very useful; I will not give proofs for all of them, but

Proposition 5.2.

(i) $T_{\mathbb{C}} = 0 \iff T = 0$. (This is obvious from the definition of $T_{\mathbb{C}}$.)

(ii) $(T + S)_{\mathbb{C}} = T_{\mathbb{C}} + S_{\mathbb{C}}$. (This is also obvious from the definition.)

(iii) For any real polynomial $p(x) \in \mathcal{P}(\mathbb{R})$, we have $p(T_{\mathbb{C}}) = \big(p(T)\big)_{\mathbb{C}}$. For example, taking $p(x) = x^2$ this means that $(T_{\mathbb{C}})^2 = (T^2)_{\mathbb{C}}$, which we can check as follows:

$$T_{\mathbb{C}}^2(v + iw)) = T_{\mathbb{C}}\big(T(v) + iT(w)\big) = T^2(v) + iT^2(w) = (T^2)_{\mathbb{C}}(v + iw).$$

(iv) $m_{T_{\mathbb{C}}}(x) = m_T(x)$. (This is not as easy as I said in class, but it is true.)

(v) $\det(T_{\mathbb{C}}) = \det(T)$. (This is pretty easy just using the definition of determinant.)

(vi) $\chi_{T_{\mathbb{C}}}(x) = \chi_T(x)$. (This follows from (v).)

Proof of Cayley–Hamilton over $\mathbb{R}$. We can now immediately deduce the Cayley–Hamilton theorem for real operators from the theorem for complex operators. Let V be a finite-dimensional real vector space, and let $T \in \mathcal{L}(V)$ be a real operator on V. Let $p(x) \in \mathcal{P}(\mathbb{R})$ be the characteristic polynomial of T:

$$p(x) = \chi_T(x)$$

Our goal is to prove that $p(T)$, or in other words $\chi_T(T)$, is 0.

Let $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ be the complexification of the operator T. By Proposition 5.2(vi), $p(x)$ is also the characteristic polynomial of $T_{\mathbb{C}}$:

$$p(x) = \chi_{T_{\mathbb{C}}}(x) = \chi_T(x)$$

Therefore by the Cayley–Hamilton theorem for the complex operator $T_{\mathbb{C}}$, we know that

$$p(T_{\mathbb{C}}) = 0.$$

By Proposition 5.2(iii), this is equivalent to

$$\big(p(T)\big)_{\mathbb{C}} = 0.$$

But then Proposition 5.2(i) implies that $p(T) = 0$, as desired. $\square$