

Mathematics 595 (CAP/TRA)
Introduction to Linear Analysis
Fall 2005

1 Linear Algebra

1.1 Vector spaces

Definition 1.1.1. Let \mathbb{F} be a field. A *vector space over \mathbb{F}* is a set V of elements, called *vectors*, equipped with two operations $+$: $V \times V \rightarrow V$ (addition) and \cdot : $\mathbb{F} \times V \rightarrow V$ (scalar multiplication) such that $(V, +)$ is an abelian group, and the following compatibility relations hold:

1. (distributivity) $a, b \in \mathbb{F}, v, w \in V \Rightarrow a(v + w) = av + aw, (a + b)v = av + bv,$
2. (associativity of multiplication) $a(bv) = (ab)v,$
3. (identity) $1v = v.$

Here 1 denotes the unit element of \mathbb{F} . Notice that we typically omit the notation \cdot for scalar multiplication, writing $av = a \cdot v$.

Examples 1.1.2. 1. n -dimensional Euclidean space \mathbb{R}^n is a vector space over \mathbb{R} .

2. The space $C[0, 1]$ of continuous functions on the interval $[0, 1]$ is a vector space over \mathbb{R} .
3. Let $k = 0, 1, 2, \dots$. The space $C^k[0, 1]$ of functions f on the interval $[0, 1]$ whose k th derivative $f^{(k)}$ exists and is continuous is a vector space over \mathbb{R} . Similarly, the space $C^\infty[0, 1]$ of functions on $[0, 1]$ for which $f^{(k)}$ exists for all k , is a vector space over \mathbb{R} .
4. \mathbb{R} is a vector space over \mathbb{Q} (see Exercise 1.1.17).
5. This example requires some basic fluency in abstract algebra. Let p be a prime and let K be a finite field of characteristic p . Then K is a vector space over \mathbb{Z}_p .
6. Let V be a vector space over \mathbb{F} , and let $W \subset V$ be closed under addition and scalar multiplication: $v, w \in W, a \in \mathbb{F} \Rightarrow v + w \in W, av \in W$. Then W is a vector space over \mathbb{F} (called a (*vector*) *subspace* of V).
7. Let W be a vector subspace of V , both over a field \mathbb{F} . Define a relation on V as follows: $v \sim v'$ iff $v - v' \in W$. Then \sim is an equivalence relation on V , and the collection of equivalence classes V/\sim is a vector space over \mathbb{F} . This space is called the *quotient space* of V by W , and is written V/W . We typically write elements of V/W in the form $v + W, v \in V$, with the understanding that $v + W = v' + W$ iff $v \sim v'$, i.e., $v - v' \in W$.

8. Let V and W be vector spaces over \mathbb{F} . The product space $V \times W$ equipped with the operations $(v, w) + (v', w') = (v + v', w + w')$ and $a(v, w) = (av, aw)$ is a vector space over \mathbb{F} . It is called the *direct sum* of V and W , and is written $V \oplus W$.

Exercise 1.1.3. Verify the claims in the various parts of Example 1.1.2.

Throughout this course, we will work almost entirely with either the field \mathbb{R} of real numbers or the field \mathbb{C} of complex numbers. From now on we omit the phrase “over \mathbb{F} ”.

Let S be any set. A set $R \subset S$ is called *cofinite* if $S \setminus R$ is a finite set. A function $\lambda : S \rightarrow \mathbb{F}$ is called *essentially zero* if $\{s : \lambda(s) \neq 0\}$ is a cofinite subset of S .

Definition 1.1.4. Let V be a vector space. A set $S \subset V$ is said to be *linearly independent* if whenever $\lambda : S \rightarrow \mathbb{F}$ is essentially zero, and

$$\sum_{s \in S} \lambda(s)s = 0,$$

then λ is identically zero. A set $S \subset V$ is *linearly dependent* if it is not linearly independent.

Notice that the sum $\sum_{s \in S} \lambda(s)s$ only involves a finite number of nonzero terms, hence is well-defined.

Definition 1.1.5. Let V be a vector space, and let $S \subset V$. A *linear combination* of elements of S is an element of V which may be written in the form

$$\sum_{s \in S} \lambda(s)s$$

for some essentially zero function $\lambda : S \rightarrow \mathbb{F}$. The *span of S* is the set $\text{span}(S)$ of linear combinations of elements of S .

Definition 1.1.6. A set $B \subset V$ is a *basis* if B is linearly independent and $\text{span}(B) = V$.

Theorem 1.1.7. *Every vector space admits a basis.*

Proof. We will use Zorn’s Lemma: every partially ordered set P , such that all totally ordered subsets of P admit upper bounds in P , has maximal elements.

Let V be a vector space, and let \mathcal{L} be the collection of all linearly independent subsets of V , partially ordered by inclusion. Let $(S_i)_{i \in I}$ be a chain in \mathcal{L} , i.e. a totally ordered subset. We claim that $\bigcup_i S_i$ is in \mathcal{L} , i.e., $\bigcup_i S_i$ is linearly independent. If $\lambda : \bigcup_i S_i \rightarrow \mathbb{F}$ is essentially zero and $\sum_{s \in \bigcup_i S_i} \lambda(s)s = 0$, then $\lambda(s) \neq 0$ only for finitely many elements s_{i_1}, \dots, s_{i_m} in $\bigcup_i S_i$. Then there exists an index k so that $s_{i_j} \in S_k$ for all $j = 1, \dots, m$. Since S_k is linearly independent, we conclude that λ is identically zero as desired.

Zorn’s lemma guarantees the existence of a maximal element B in \mathcal{L} . We claim that $V = \text{span}(B)$. To do this, we will show that $v \notin \text{span}(B) \Rightarrow B \cup \{v\}$ is linearly

independent. Suppose that $v \notin \text{span}(B)$ and there exists an essentially zero function $\lambda : B \cup \{v\} \rightarrow \mathbb{F}$ such that

$$\lambda(v)v + \sum_{w \in B} \lambda(w)w = 0.$$

If $\lambda(v) \neq 0$, then $v = \sum_{w \in B} (-\frac{\lambda(w)}{\lambda(v)})w$, which contradicts the fact that $v \notin \text{span}(B)$. Thus $\lambda(v) = 0$. But then

$$\sum_{w \in B} \lambda(w)w = 0;$$

since B is linearly independent we find $\lambda \equiv 0$ and so $B \cup \{v\}$ is linearly independent. \square

Corollary 1.1.8. *Any linearly independent set S in a vector space V can be extended to a basis.*

Proof. Repeat the proof of the theorem, replacing \mathcal{L} with the collection of all linearly independent subsets of V which contain S . \square

Theorem 1.1.9. *Any two bases for a vector space V have the same cardinality.*

Proof. We consider the finite cardinality case. Suppose that V has a basis $B = \{v_1, \dots, v_m\}$ consisting of m elements. It suffices to prove that every collection of $m + 1$ elements in V is linearly dependent. (Why?) Let $C = \{w_1, \dots, w_{m+1}\}$ be such a collection. Consider the expression

$$\sigma = c_1w_1 + \dots + c_{m+1}w_{m+1}.$$

Since each w_j may be written as a linear combination of the elements of B , σ may be written as a linear combination of v_1, \dots, v_m . Setting $\sigma = 0$ generates a system of m linear homogeneous equations in the $m + 1$ variables c_1, \dots, c_{m+1} , which necessarily has a nontrivial solution. But such a nontrivial solution is clearly equivalent to a linear dependence among the vectors w_1, \dots, w_{m+1} .

The infinite cardinality case can be treated by transfinite induction. We omit the proof. \square

Definition 1.1.10. Let V be a vector space. The number of elements in a basis of V is called the *dimension* of V and written $\dim V$. We say that V is *finite dimensional* if $\dim V < \infty$.

Example 1.1.11. The *standard basis* for \mathbb{R}^n is the collection of n vectors e_1, \dots, e_n , where the coordinates of e_j are equal to 0 in all entries except for the j th entry, where the coordinate is equal to 1.

Exercise 1.1.12. Let V be a vector space of dimension n , and let S be a set of n linearly independent elements of V . Then S is a basis for V .

Exercise 1.1.13. Let $V = W_1 \oplus W_2$ be finite dimensional. Then $\dim V = \dim W_1 + \dim W_2$.

Exercise 1.1.14. Let W be a subspace of a finite dimensional vector space V . Then $\dim V = \dim W + \dim V/W$. (This result will also follow as an easy consequence of Theorem 1.2.21.)

Let B be a basis for a vector space V and let $S \subset V$. For each $s \in S$, there exist scalars λ_{sb} such that

$$s = \sum_{b \in B} \lambda_{sb} b.$$

Definition 1.1.15. The matrix $M_{SB} := (\lambda_{sb})_{s \in S, b \in B}$ is called the *transition matrix* of S with respect to B .

Lemma 1.1.16. Let B and S be bases for V with transition matrices $M_{SB} = (\lambda_{sb})$ and $M_{BS} = (\mu_{bs})$. Then

$$I = M_{BS}M_{SB} = M_{SB}M_{BS}.$$

Note that M_{SB} and M_{BS} may be infinite matrices, in which case the meaning of the matrix multiplications may not be immediately clear. However, the fact that the coefficients λ_{sb} define essentially zero functions of s and b , provided the remaining variable is fixed, ensures that the notion of matrix multiplication is well-defined for such matrices.

Proof. For $b \in B$, we have

$$b = \sum_{s \in S} \mu_{bs} s = \sum_{s \in S} \sum_{b' \in B} \mu_{bs} \lambda_{sb'} b' = \sum_{b' \in B} \left(\sum_{s \in S} \mu_{bs} \lambda_{sb'} \right) b'.$$

By the linear independence of B , we find

$$\sum_{s \in S} \mu_{bs} \lambda_{sb'} = \delta_{bb'},$$

where $\delta_{bb'}$ denotes the Kronecker delta function: $\delta_{bb'} = 0$ if $b \neq b'$ and $\delta_{bb} = 1$. In other words,

$$M_{BS}M_{SB} = I.$$

The other direction is similar. □

Exercise 1.1.17. What is the cardinality of a basis for \mathbb{R} over \mathbb{Q} ?

1.2 Linear Maps

Definition 1.2.18. Let V and W be vector spaces. A map $T : V \rightarrow W$ is called *linear* if

$$T(av + bw) = aT(v) + bT(w)$$

for all $v, w \in V$ and $a, b \in \mathbb{F}$. We use the terms *linear map* and *linear transformation* interchangeably. An *isomorphism* is a bijective linear map. Two vector spaces V and W are called *isomorphic* (written $V \simeq W$) if there exists an isomorphism $T : V \rightarrow W$.

A few easy observations:

1. $T(V)$ is a subspace of W .
2. If W_1 is a subspace of W , then $T^{-1}(W_1) := \{v \in V : T(v) \in W_1\}$ is a subspace of V .
3. If $S : U \rightarrow V$ and $T : V \rightarrow W$ are linear, then $T \circ S : U \rightarrow W$ is linear.
4. If T is a bijective linear map, then T^{-1} is linear.
5. If T is surjective, then $\dim V \geq \dim W$. If T is injective, then $\dim V \leq \dim W$. In particular, isomorphic vector spaces have the same dimension.

Examples 1.2.19. 1. $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, A an $m \times n$ real matrix, $T(v) = Av$.

2. $V = C^k[0, 1]$, $W = C^{k-1}[0, 1]$, $T(f) = f'$.

3. W a subspace of a vector space V , $T : V \rightarrow V/W$ given by $T(v) = v + W$. This example will be important for us in later work. The map $T : V \rightarrow V/W$ is called a *quotient map*.

Definition 1.2.20. The *kernel* of a linear map $T : V \rightarrow W$ is

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

The *range* of T is $\text{range}(T) = T(V)$.

Theorem 1.2.21 (Rank–Nullity Theorem). Let $T : V \rightarrow W$ be a linear map. Then

$$V \simeq \ker(T) \oplus \text{range}(T)$$

and

$$\dim \ker(T) + \dim \text{range}(T) = \dim V.$$

Corollary 1.2.22. Let W be a subspace of a vector space V . Then $V \simeq W \oplus V/W$ and $\dim V/W = \dim V - \dim W$.

Proof. Apply Theorem 1.2.21 to the linear map $Q : V \rightarrow V/W$ given by $Q(v) = v + W$. □

Corollary 1.2.23. *Let V be a finite dimensional vector space and let $T : V \rightarrow V$ be linear. Then T is injective if and only if T is surjective.*

Proof. T is injective if and only if $\dim \ker(T) = 0$. T is surjective if and only if $\dim \text{range}(T) = \dim V$. \square

Exercise 1.2.24. Show that Corollary 1.2.23 is false for infinite dimensional vector spaces; a linear self-map of such a space can be injective without being surjective, and vice versa.

Proof of Theorem 1.2.21. Let $S_1 = (v_i)_{i \in I}$ be a basis for $\ker(T)$, and let $T(S_2)$, $S_2 = (w_j)_{j \in J}$, be a basis for $\text{range}(T)$. We claim that $S_1 \cup S_2$ is a basis for V . This suffices to complete the proof, since we may then exhibit an isomorphism between V and $\ker(T) \oplus \text{range}(T)$ by sending v_i to $(v_i, 0)$ and w_j to $(0, T(w_j))$.

First, we show that $S_1 \cup S_2$ is linearly independent. Let $\lambda : S_1 \cup S_2 \rightarrow \mathbb{F}$ be essentially zero, with

$$0 = \sum_{v \in S_1 \cup S_2} \lambda(v)v = \sum_i \lambda(v_i)v_i + \sum_j \lambda(w_j)w_j.$$

Then

$$0 = T(0) = T\left(\sum_i \lambda(v_i)v_i + \sum_j \lambda(w_j)w_j\right) = \sum_j \lambda(w_j)T(w_j)$$

since $v_i \in \ker(T)$. Since $T(S_2)$ is a basis for $\text{range}(T)$, we see that $\lambda(w_j) = 0$ for all $j \in J$. Thus

$$0 = \sum_i \lambda(v_i)v_i;$$

since S_1 is a basis for $\ker(T)$ we conclude that $\lambda(v_i) = 0$ for all $i \in I$. Thus $\lambda \equiv 0$ and we have shown that $S_1 \cup S_2$ is linearly independent.

Next, we show that $S_1 \cup S_2$ spans all of V . Given any $v \in V$, we may write $T(v)$ as a linear combination of the elements of $T(S_2)$:

$$T(v) = \sum_j \mu(w_j)T(w_j)$$

for a suitable essentially zero function $\mu : S_2 \rightarrow \mathbb{F}$. Then $v - \sum_j \mu(w_j)w_j \in \ker(T)$ whence

$$v - \sum_j \mu(w_j)w_j = \sum_i \lambda(v_i)v_i$$

for some essentially zero function $\lambda : S_1 \rightarrow \mathbb{F}$. This exhibits v as a linear combination of $S_1 \cup S_2$, as desired. \square

Any linear map may be factored through its kernel by an injective linear map.

Proposition 1.2.25. *Let $T : V \rightarrow W$ be a linear map, and let $Q : V \rightarrow V/\ker(T)$ be the quotient map $Q(v) = v + \ker(T)$. Then there exists a unique injective linear map $T' : V/\ker(T) \rightarrow W$ such that $T' \circ Q = T$.*

Proof. Define $T'(v + \ker(T)) = T(v)$. This is well-defined: if $v' + \ker(T) = v + \ker(T)$ then $v' - v \in \ker(T)$ so $T(v' - v) = 0$, i.e., $T(v') = T(v)$. It is obviously linear. To see that it is injective, suppose that $T'(v + \ker(T)) = T'(v' + \ker(T))$. Then $T(v) = T(v')$ so $v - v' \in \ker(T)$ and $v + \ker(T) = v' + \ker(T)$. \square

The collection of all linear maps from V to W is denoted by $L(V, W)$. Of particular interest is the space $L(V, V)$ of linear maps of V into itself. Note that any two such maps can be added and multiplied (composition), and multiplied by a scalar. Thus $L(V, V)$ is an example of an *algebra*. It is typically not commutative, since $S \circ T$ need not equal $T \circ S$. Furthermore, $L(V, V)$ typically contains *zerodivisors*, elements $S, T \neq 0$ such that $S \circ T = 0$. For example, choose S and T so that $\text{range}(S) \subset \ker(T)$.

The set of invertible elements in $L(V, V)$ forms a group. The isomorphism type of this group depends only on the dimension of V , and on the scalar field \mathbb{F} . It is called the *general linear group* of dimension $n = \dim V$ over \mathbb{F} , and denoted by $\text{GL}(n, \mathbb{F})$.

Given $S \in \text{GL}(n, \mathbb{F})$, the transformation $T \mapsto STS^{-1}$ is an invertible linear map of $L(\mathbb{F}^n, \mathbb{F}^n)$, i.e., an element of $\text{GL}(N, \mathbb{F})$, where $N = \dim L(\mathbb{F}^n, \mathbb{F}^n)$. It is called a *similarity transformation*, and the maps T and STS^{-1} are said to be *similar* to each other.

Exercise 1.2.26. Verify that similarity transformations of $V = \mathbb{F}^n$ are elements of $\text{GL}(N, \mathbb{F})$, $N = \dim L(V, V)$. What is the value of $\dim L(V, V)$?

Linear maps from V to the base field \mathbb{F} are called *linear functionals* on V . The space $L(V, \mathbb{F})$ of linear functionals on V is also known as the *dual space* of V , and is sometimes denoted V^* .

Exercise 1.2.27. Let V be a finite dimensional vector space of dimension n , with basis v_1, \dots, v_n .

1. Show that every element $v \in V$ can be written uniquely in the form $v = \sum_{i=1}^n \lambda_i v_i$.
2. Show that the functions $k_i : V \rightarrow \mathbb{F}$ given by $k_i(v) = \lambda_i$, where λ_i is as in the previous part, are linear functionals on V .
3. Show that every linear functional T on V can be written in the form $T = \sum_{i=1}^n a_i k_i$, for some $a_i \in \mathbb{F}$.
4. Show that the mapping $T \mapsto \sum_{i=1}^n a_i v_i$ defines a (non-canonical) isomorphism between V^* and V . Conclude that $\dim V^* = \dim V$.

Exercise 1.2.28. Let P be the vector space of polynomials over \mathbb{C} with degree strictly less than n , and let $a_1, \dots, a_n \in \mathbb{C}$ and $c_1, \dots, c_n \in \mathbb{C}$ be arbitrary. Show that there is an element $p \in P$ so that $p(a_i) = c_i$ for all $i = 1, \dots, n$.

Matrix representations of linear maps: If $T : V \rightarrow W$ is a linear map and C , resp. B , are bases for V resp. W , then each element $c \in C$ can be written uniquely in the form

$$c = \sum_{b \in B} \lambda_{cb} b.$$

In the finite-dimensional case, the matrix $M = [T]_{C,B} = (\lambda_{cb})_{c \in C, b \in B}$ is called the *matrix representation for T* with respect to the bases C and B .

Lemma 1.2.29. *Let A and B be bases for V , let C and D be bases for W , and let $T : V \rightarrow W$ be a linear map. Then*

$$[T]_{D,A} = M_{DC} [T]_{C,B} M_{BA}.$$

Corollary 1.2.30. *If A and B are bases for V and $T \in L(V, V)$, then*

$$[T]_{A,A} = M_{AB} [T]_{B,B} M_{BA} = M_{AB} [T]_{B,B} M_{AB}^{-1}.$$

In particular, the matrices $[T]_{A,A}$ and $[T]_{B,B}$ are similar.

1.3 Geometry of linear transformations of finite-dimensional vector spaces

Instead of beginning with the algebraic definitions for determinant and trace, we will take a geometric perspective. This is arguably more enlightening, as it indicates the underlying physical meaning for these quantities.

We assume that the notion of volume of open sets in \mathbb{R}^n is understood. In fact, we will only need to work with the volumes of simplexes, which admit a simple recursive definition (see Definition 1.3.34). The definition relies on the geometric structure of \mathbb{R}^n , specifically the notions of perpendicularity (orthogonality) and distance, which we take as known.

Definition 1.3.31. An *oriented simplex* in \mathbb{R}^n is the convex hull of a set of $n + 1$ vertices. More precisely, the simplex spanned by points $a_0, \dots, a_n \in \mathbb{R}^n$ is

$$S = [a_0, \dots, a_n] = \{x \in \mathbb{R}^n : x = \sum_{i=0}^n \lambda_i a_i \text{ for some } \lambda_i \geq 0 \text{ with } \sum_i \lambda_i = 1\}.$$

The order of the vertices matters in this notation, i.e., $[a_0, a_1, \dots, a_n]$ and $[a_1, a_0, \dots, a_n]$ are different as simplices (although they define the same set in \mathbb{R}^n).

An oriented simplex S is called *degenerate* if it lies in an $(n - 1)$ -dimensional subspace of \mathbb{R}^n .

Exercise 1.3.32. Show that $[a_0, \dots, a_n]$ is degenerate if and only if the vectors $a_1 - a_0, \dots, a_n - a_0$ are linearly dependent.

Our considerations will be invariant under rigid translations of \mathbb{R}^n , so we may (and will) from now on assume that $a_0 = 0$, since we may identify $[a_0, \dots, a_n]$ with $[0, a_1 - a_0, \dots, a_n - a_0]$.

Next, we would like to assign an orientation $O(S)$ to each nondegenerate oriented simplex S . We will classify all such simplexes as either *positively oriented* ($O(S) = +1$) or *negatively oriented* ($O(S) = -1$), so that the following requirements are satisfied:

1. (anti-symmetry) The orientation of S is reversed if any two vertices a_i and a_j of S are interchanged.
2. (continuity) The orientation $O(S(t))$ is constant when $S(t) = [0, a_1(t), \dots, a_n(t)]$ is a continuous family of nondegenerate simplexes.
3. (normalization) The orientation of the standard n -dimensional simplex $S_0^n = [0, e_1, \dots, e_n]$ is $+1$.

Here (e_j) denotes the standard basis for \mathbb{R}^n , see Example 1.1.11.

Proposition 1.3.33. *There is a unique notion of orientation for simplexes in \mathbb{R}^n satisfying these three requirements.*

Proof. Consider those simplexes $S = [a_0, \dots, a_n]$ which may be deformed by a continuous family of nondegenerate simplexes to the standard simplex. That is, there exist n continuous functions $a_j : [0, 1] \rightarrow \mathbb{R}^n$, $j = 1, \dots, n$, so that (i) $a_j(0) = e_j$ for

all j , (ii) $a_j(1) = a_j$ for all j , and (iii) $S(t) = [0, a_1(t), \dots, a_n(t)]$ is nondegenerate for all $0 \leq t \leq 1$. We call each of these simplexes *positively oriented* and assign it the orientation $O(S) = +1$. Otherwise we call S *negatively oriented* and assign it the orientation $O(S) = -1$. It is straightforward to check that all three requirements are satisfied in this case. \square

Definition 1.3.34. The n -volume of a simplex $S = [0, a_1, \dots, a_n] \subset \mathbb{R}^n$ is

$$\text{Vol}_n(S) = \frac{1}{n} \text{Vol}_{n-1}(B)h,$$

where $B = [0, a_1, \dots, \widehat{a}_j, a_n]$ is a base of S (the notation means that the point a_j is omitted from the list a_1, \dots, a_n) and h is the altitude of S with respect to B , i.e., h is the distance from a_j to the $(n-1)$ -dimensional hyperplane Π containing B . (To begin the induction, we let Vol_1 be the standard length measure for intervals on the real line \mathbb{R}^1 .) To simplify the notation, we drop the subscripts, writing $\text{Vol} = \text{Vol}_n$ for every n . Notice that $\text{Vol}(S) = 0$ if and only if S is degenerate.

The *signed volume* of S is

$$\text{signVol}(S) = O(S) \text{Vol}(S).$$

Proposition 1.3.35. *The signed volume satisfies the following properties:*

1. $\text{signVol}(S) = 0$ if and only if S is degenerate.
2. $\text{signVol}([0, a_1, \dots, a_n])$ is a linear function of the vertices a_1, \dots, a_n .

The first claim is obvious by the definition of $O(S)$ and $\text{Vol}(S)$. To prove the second claim, it is enough to consider $\text{signVol}([0, a_1, \dots, a_n])$ as a function of a_n (by anti-symmetry). From the definitions we have

$$\text{signVol}([0, a_1, \dots, a_n]) = \frac{1}{n} \text{Vol}_{n-1}(B) \cdot O(S)h.$$

The expression $O(S)h$ represents the *signed distance* from a_n to Π . This is nothing more than a particular coordinate function, if coordinates in \mathbb{R}^n are chosen so that one axis is perpendicular to Π and the remaining $n-1$ axes are chosen to lie in Π . Exercise 1.2.27 (2) shows that this is a linear function.

We now come to the definition of determinant.

Definition 1.3.36. Let A be an $n \times n$ matrix whose columns are given by the vectors a_1, \dots, a_n . The *determinant* of A is equal to

$$\det(A) = n! \cdot \text{signVol}([0, a_1, \dots, a_n]).$$

An immediate corollary of Proposition 1.3.35 is

Proposition 1.3.37. *The determinant satisfies the following properties:*

1. (*antisymmetry*) $\det(A)$ is an alternating function of the columns of A ; its sign changes when two columns of A are interchanged. In particular, $\det(A) = 0$ if two columns of A are identical.

2. (multilinearity) $\det(A)$ is a linear function of the columns of A .
3. (normalization condition) $\det(I) = +1$, where I is the $n \times n$ identity matrix, with columns e_1, \dots, e_n (in that order).
4. $\det(A) = 0$ if and only if the columns of A are linearly dependent.
5. $\det(A)$ is unchanged if a multiple of one column of A is added into another column.

In fact, the determinant is the unique function on $n \times n$ matrices satisfying parts 1, 2 and 3 of Proposition 1.3.37.

Part 1 is immediate from the properties of $O(S)$ and $\text{Vol}(S)$. Parts 2 and 4 follow from Proposition 1.3.35 and Exercise 1.3.32. To verify part 3, we first compute

$$\text{Vol}(S_0^n) = \frac{1}{n} \text{Vol}(S_0^{n-1})$$

so

$$\text{Vol}(S_0^n) = \frac{1}{n!}$$

and

$$\det(I) = n! \cdot \text{signVol}(S_0^n) = n! O(S_0^n) \text{Vol}(S_0^n) = (n!)(+1)\left(\frac{1}{n!}\right) = 1.$$

Finally, part 5 follows from antisymmetry and multilinearity.

The determinant is a multiplicative function:

Theorem 1.3.38. $\det(AB) = \det(A) \det(B)$ for $n \times n$ matrices A, B .

Exercise 1.3.39. Use the uniqueness assertion for the determinant function to prove Theorem 1.3.38. Hint: In the case $\det A \neq 0$, consider the function

$$C(B) = \frac{\det(AB)}{\det A}.$$

Show that C satisfies Proposition 1.3.37 parts 1, 2 and 3. Note that, upon writing out $B = (b_1, \dots, b_n)$ in columns, we have

$$C(b_1, \dots, b_n) = \frac{\det(Ab_1, \dots, Ab_n)}{\det A}.$$

For the case $\det A = 0$, use a continuity argument.

Definition 1.3.40. The ij th minor of an $n \times n$ matrix $A = (a_{ij})$ is the $(n-1) \times (n-1)$ matrix \tilde{A}_{ij} obtained by deleting the i th row and j th column of A . The adjoint matrix of A is the matrix $\text{adj } A = (b_{ij})$, where

$$b_{ij} = (-1)^{i+j} \det \tilde{A}_{ji}.$$

Lemma 1.3.41. (1) For each j , $\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det \tilde{A}_{ij}$.
 (2) $A \text{ adj } A = \det A \cdot I$.

Proof. To prove (1) we begin with the special case

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{A}_{11} \end{pmatrix},$$

where the zeros denote rows (and columns) of zero entries. The map $\tilde{A}_{11} \mapsto \det A$ clearly satisfies parts 1, 2 and 3 of Proposition 1.3.37, whence $\det A = \det \tilde{A}_{11}$ in this case. Next, the case

$$A = \begin{pmatrix} 1 & * \\ 0 & \tilde{A}_{11} \end{pmatrix}$$

(where $*$ denotes a row of unspecified entries) follows from part 5. Finally, the general case follows by antisymmetry and multilinearity of the determinant. We leave the full details as an exercise to the reader.

To prove (2), we compute the entries of $C = A \operatorname{adj} A = (c_{ij})$:

$$\begin{aligned} c_{ii} &= \sum_{k=1}^n (-1)^{i+k} a_{ik} (\operatorname{adj} A)_{ki} \\ &= \sum_{k=1}^n (-1)^{i+k} a_{ik} \det \tilde{A}_{ik} = \det A \end{aligned}$$

by (1), while for $i \neq j$,

$$\begin{aligned} c_{ij} &= \sum_{k=1}^n (-1)^{j+k} a_{ik} (\operatorname{adj} A)_{kj} \\ &= \sum_{k=1}^n (-1)^{j+k} a_{ik} \det \tilde{A}_{jk} = \det A'_{ij} \end{aligned}$$

where A'_{ij} is the $n \times n$ matrix whose columns are the same as those of A , except that the j th column is omitted and replaced by a copy of the i th column. By antisymmetry, $\det A'_{ij} = 0$ so

$$c_{ij} = \det A \cdot \delta_{ij}$$

and the proof is complete. \square

Part (1) of Lemma 1.3.41 is the well-known *Laplace expansion* for the determinant.

Definition 1.3.42. A matrix A is *invertible* if $\det A \neq 0$. The *inverse* of an invertible matrix A is

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

Lemma 1.3.41(2) yields the standard properties of the inverse:

$$A \cdot A^{-1} = A^{-1} \cdot A = I.$$

Proposition 1.3.43. *The determinant of a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is independent of the choice of basis of \mathbb{R}^n .*

Proof. Changing the basis of \mathbb{R}^n corresponds to performing a similarity transformation $T \mapsto STS^{-1}$. Denoting by $A \mapsto BAB^{-1}$ the matrix representation of this transformation with respect to the standard basis of \mathbb{R}^n , we have

$$\det(BAB^{-1}) = \det(B) \det(A) \det(B^{-1}) = \det(A).$$

□

Because of this proposition, we may write $\det(T)$ for $T \in L(\mathbb{R}^n, \mathbb{R}^n)$ to denote the determinant of any matrix representation $[T]_{B,B}$ for T . Moreover, the notion of invertibility is well-defined for elements of $L(\mathbb{R}^n, \mathbb{R}^n)$.

Corollary 1.3.44. *Let T be a linear map of \mathbb{R}^n to \mathbb{R}^n . Then*

$$|\det(T)| = \frac{\text{Vol}(T(S))}{\text{Vol}(S)}$$

for any nondegenerate simplex S in \mathbb{R}^n .

This corollary provides the “true” geometric interpretation of the determinant: it measures the distortion of volume induced by the linear map.

Definition 1.3.45. The *trace* of an $n \times n$ matrix A with columns a_1, \dots, a_n is

$$\text{tr}(A) = \sum_{i=1}^n a_i \cdot e_i.$$

Here $v \cdot w$ denotes the usual Euclidean dot product of vectors $v, w \in \mathbb{R}^n$. If $a_i = (a_{i,1}, \dots, a_{i,n})$ then $\text{tr}(A) = \sum_{i=1}^n a_{ii}$.

Exercise 1.3.46. 1. Show that trace is a linear function on matrices: $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$ and $\text{tr}(kA) = k \text{tr}(A)$.

2. Show that $\text{tr}(AB) = \text{tr}(BA)$ for any $n \times n$ matrices A and B .

Proposition 1.3.47. *The trace of a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is independent of the choice of basis of \mathbb{R}^n .*

Proof. Changing the basis of \mathbb{R}^n corresponds to performing a similarity transformation $T \mapsto STS^{-1}$. If $A \mapsto BAB^{-1}$ denotes the matrix representation of this transformation with respect to the standard basis of \mathbb{R}^n , then

$$\text{tr}(BAB^{-1}) = \text{tr}(AB^{-1}B) = \text{tr}(A)$$

by Exercise 1.3.46 (2). □

Because of this proposition, we may write $\text{tr}(T)$ for $T \in L(\mathbb{R}^n, \mathbb{R}^n)$ to denote the trace of any matrix representation for T .

1.4 Spectral theory of finite-dimensional vector spaces and Jordan canonical form

Definition 1.4.48. Let V be a vector space and let $T \in L(V, V)$. An element $\lambda \in \mathbb{F}$ is called an *eigenvalue* of T if there exists a nonzero vector $v \in V$ (called an *eigenvector* for λ) so that $Tv = \lambda v$.

Equivalently, if we denote by I the identity element in $L(V, V)$, then λ is an eigenvalue of T if and only if

$$\ker(\lambda I - T) \neq 0.$$

The set of eigenvalues of T is called the *spectrum* of T , and is written $\text{Spec } T$.

Exercise 1.4.49. Let λ be an eigenvalue for a linear transformation T on a vector space V . The set of eigenvectors associated with λ is a vector subspace of V .

For the rest of this section, we assume that V is a finite-dimensional vector space, defined over an algebraically closed field \mathbb{F} (typically \mathbb{C}).

Definition 1.4.50. The *characteristic polynomial* for a linear transformation T on V is

$$p_T(\lambda) = \det(\lambda I - T).$$

Here λ denotes an *indeterminate variable*, and the computation of the determinant is understood in terms of the formulas from the previous section in the field $\mathbb{F}[[\lambda]]$ of rational functions in λ with coefficients from \mathbb{F} .

The Cayley–Hamilton theorem states that A satisfies its own characteristic polynomial.

Theorem 1.4.51 (Cayley–Hamilton). *For any linear transformation T on a finite-dimensional vector space V , the identity $p_T(T) = 0$ holds.*

Proof. By the invariance of determinant with respect to basis, we may choose a basis B for V and study the matrix representation $A = [T]_{B,B}$. By Lemma 1.3.41(2),

$$(\lambda I - A) \text{adj}(\lambda I - A) = \det(\lambda I - A)I = p_A(\lambda)I. \quad (1.4.52)$$

Set

$$\text{adj}(\lambda I - A) := \sum_{k=0}^n B_k \lambda^k,$$

where $B_n = 0$, and set

$$p_A(\lambda) := \sum_{k=0}^n c_k \lambda^k,$$

where $c_n = 1$. Then (1.4.52) gives

$$\begin{aligned} \sum_{k=0}^n c_k \lambda^k I &= (\lambda I - A) \text{adj}(\lambda I - A) = (\lambda I - A) \sum_{k=0}^n B_k \lambda^k \\ &= -AB_0 + \sum_{k=1}^n (B_{k-1} - AB_k) \lambda^k. \end{aligned}$$

Thus $B_{n-1} = I$, $-AB_0 = c_0I$, and

$$B_{k-1} - AB_k = c_kI, \quad \text{for } k = 1, \dots, n-1.$$

Multiplying by A^k gives

$$A^k B_{k-1} - A^{k+1} B_k = c_k A^k, \quad \text{for } k = 1, \dots, n-1.$$

Finally,

$$p_A(A) = \sum_{k=0}^n c_k A^k = A^n + \sum_{k=1}^{n-1} (A^k B_{k-1} - A^{k+1} B_k) - AB_0 = 0.$$

□

By the fundamental theorem of algebra, each $n \times n$ matrix has n complex eigenvalues (counted with multiplicity).

Exercise 1.4.53. Show that $p_T(\lambda) = \lambda^n - (\text{tr } T)\lambda^{n-1} + \dots + (-1)^n \det T$ for any T . Deduce that the sum of the eigenvalues of T is $\text{tr } T$ and that the product of the eigenvalues is $\det T$. (Eigenvalues must be counted according to multiplicity.)

The set of polynomials in $\mathbb{F}[\lambda]$ which vanish on T forms an ideal \mathcal{I} (\mathcal{I} is closed under addition and closed under multiplication by elements of $\mathbb{F}[\lambda]$.) Since $\mathbb{F}[\lambda]$ is a principal ideal domain, there exists an element $m_T \in \mathbb{F}[\lambda]$ so that $\mathcal{I} = (m_T)$, i.e., every element of \mathcal{I} is a multiple of m_T . We can and do assume that m_T is monic, and we call it the *minimal polynomial* for T . Obviously p_T is divisible by m_T .

In the principal theorem of this section (Theorem 1.4.57) we give, among other things, a geometric description of the characteristic and minimal polynomials. To obtain this description, we introduce a generalization of the notion of eigenvector.

Definition 1.4.54. A nonzero vector $v \in V$ is a *generalized eigenvector* of T with eigenvalue λ if $(\lambda I - T)^k v = 0$ for some positive integer k .

For a given eigenvalue λ of T , consider the increasing sequence of subspaces

$$N_1(\lambda) \subset N_2(\lambda) \subset \dots$$

where $N_k(\lambda) = \ker((\lambda I - T)^k)$. Since these are all subspaces of a finite-dimensional vector space V , they eventually stabilize: there exists an index K so that $N_K(\lambda) = N_{K+1}(\lambda) = \dots$. We denote by $d = d(\lambda)$ the smallest such index, i.e.,

$$N_{d-1}(\lambda) \subsetneq N_d(\lambda) = N_{d+1}(\lambda) = \dots$$

$d(\lambda)$ is called the *index* of λ . Thus the nonzero elements of $N_d(\lambda)$ are precisely the generalized eigenvectors associated with λ .

Lemma 1.4.55. For each j , $d(\lambda_j) \leq n = \dim V$.

Proof. Let $\lambda = \lambda_j$ and $d = d(\lambda_j)$ as above, and let $v \in N_d(\lambda) \setminus N_{d-1}(\lambda)$. To prove the result, it suffices to establish that the d vectors $v, (\lambda I - T)(v), \dots, (\lambda I - T)^{d-1}(v)$ are linearly independent. Suppose that $a_0, \dots, a_{d-1} \in \mathbb{F}$ satisfy

$$\sum_{k=0}^{d-1} a_k (\lambda I - T)^k (v) = 0. \quad (1.4.56)$$

Applying $(\lambda I - T)^{d-1}$ to both sides gives

$$a_0 (\lambda I - T)^{d-1}(v) = 0$$

so $a_0 = 0$. Returning to (1.4.56) and applying $(\lambda I - T)^{d-2}$ to both sides gives $a_1 = 0$. Continuing in this fashion, we find that $a_k = 0$ for all k . The proof is complete. \square

We now state the main theorem of this section, a combination of the classical spectral theorem for transformations of finite-dimensional vector spaces and the Jordan canonical form for matrix representations of such transformations.

Theorem 1.4.57 (Spectral theorem and Jordan canonical form). *Let $T \in L(V, V)$ be a linear transformation of an n -dimensional vector space V with distinct eigenvalues $\lambda_1, \dots, \lambda_m$. Then*

1. $V = N^1 \oplus \dots \oplus N^m$, where $N^j := N_{d(\lambda_j)}(\lambda_j)$. Moreover, $T : N^j \rightarrow N^j$ and $(\lambda_j I - T)|_{N^j}$ is nilpotent, in fact, $(\lambda_j I - T)^n|_{N^j} = 0$.

2.

$$p_T(\lambda) = \prod_j (\lambda - \lambda_j)^{\dim N^j} \quad (1.4.58)$$

and

$$m_T(\lambda) = \prod_j (\lambda - \lambda_j)^{d(\lambda_j)}. \quad (1.4.59)$$

3. For each j there exists a block diagram of generalized eigenvectors $\{v_{il}^j\}$, where $i = 1, \dots, k_j$ and $l = 1, \dots, p_i^j$, which forms a basis for the generalized eigenspace N^j . Here $p_1^j \geq p_2^j \geq p_3^j \geq \dots$, $\sum_i p_i^j = \dim N^j$ and

$$(T - \lambda_j I)(v_{il}^j) = v_{i,l-1}^j$$

for all relevant j, i and l .

4. The full collection $B = \bigcup_{j=1}^m \{v_{il}^j\}$ forms a basis for V . With respect to that (ordered) basis, $[T]_{B,B}$ is a block diagonal matrix with Jordan blocks.

A $k \times k$ matrix J is a *Jordan block* if it takes the form

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

Remark 1.4.60. (1) The Jordan canonical form of a matrix is uniquely determined by the following data:

- the values and multiplicities of the eigenvalues λ_j ,
- the block diagrams $\{v_{ii}^j\}$ associated with each λ_j , specifically, the number of rows k_j and the values

$$\{\dim N_i(\lambda_j)/N_{i-1}(\lambda_j) : j = 1, \dots, m, i = 1, \dots, k_j\}.$$

In other words, any two similar matrices have all of this data in common, and any two matrices which have all of this data in common are necessarily similar.

Exercise 1.4.61. For each j , show that λ_j is the only eigenvalue of $T|_{N^j} : N^j \rightarrow N^j$.

Examples 1.4.62. (1) Let

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

and let $T : \mathbb{C}^3 \rightarrow \mathbb{C}^3$ be the linear transformation $T(v) = Av$. The only eigenvalue is $\lambda = 1$, of multiplicity $r = 3$. Thus the characteristic polynomial $p_T(\lambda) = (\lambda - 1)^3$. A computation gives

$$A - I = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and $(A - I)^2 = 0$, so $d = d(1) = 2$ and $m_T(\lambda) = (\lambda - 1)^2$. The vectors e_1 and $e_1 + e_2 + e_3$ span the eigenspace $N_1(1)$, while $N_2(1) = \mathbb{C}^3$. Choosing a vector in $N_2(1) \setminus N_1(1)$, e.g., $v_{12} := e_1 + e_2$ we compute $(T - I)(v_{12}) =: v_{11} := e_1$. (Note: to simplify the notation, since there is only one eigenvalue to keep track of in this example, we omit it from the notation.) To complete a basis for \mathbb{C}^3 , we choose any other element of $N_1(1)$, e.g., $v_{21} = e_1 + e_2 + e_3$. Relative to the basis $B = \{v_{11}, v_{12}, v_{21}\}$, the transformation T takes the Jordan canonical form

$$[T]_{B,B} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(2) Let

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

and let $T : \mathbb{C}^5 \rightarrow \mathbb{C}^5$ be the linear transformation $T(v) = Av$. As before, the only eigenvalue is $\lambda = 1$, of multiplicity $r = 5$. Thus the characteristic polynomial

$p_T(\lambda) = (\lambda - 2)^5$. A computation gives

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

and $(A - 2I)^2 = 0$, so $d = d(1) = 2$ and $m_T(\lambda) = (\lambda - 2)^2$. The vectors e_1 , $e_2 - e_3$ and $e_4 - e_5$ span the eigenspace $N_1(2)$, while $N_2(2) = \mathbb{C}^5$. Choosing two vectors in $N_2(2) \setminus N_1(2)$, e.g., $v_{12} := e_1 + e_2$ and $v_{22} := e_1 + e_2 + e_3 - e_4$ we compute $(T - 2I)(v_{12}) =: v_{11} := e_1$ and $(T - 2I)(v_{22}) =: v_{21} := e_1 + e_2 - e_3$. To complete a basis for \mathbb{C}^5 , we choose any other element of $N_1(2)$, e.g., $v_{31} = e_1 + e_2 - e_3 + e_4 - e_5$. Relative to the basis $B = \{v_{11}, v_{12}, v_{21}, v_{22}, v_{31}\}$, the transformation T takes the Jordan canonical form

$$[T]_{B,B} = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

We now start the proof of Theorem 1.4.57.

Lemma 1.4.63. *Generalized eigenvectors associated with distinct eigenvalues are linearly independent.*

Proof. Let $\lambda_1, \dots, \lambda_m$ be distinct eigenvalues of a transformation $T : V \rightarrow V$ of an n -dimensional vector space V , and suppose that v_1, \dots, v_m are associated generalized eigenvectors. Suppose that

$$a_1 v_1 + \dots + a_m v_m = 0; \tag{1.4.64}$$

we must show that $a_j = 0$ for all j .

Let j be arbitrary. Choose $k \leq d(\lambda_j)$ so that $v_j \in N_k(\lambda_j) \setminus N_{k-1}(\lambda_j)$. Applying the operator

$$(\lambda_j I - T)^{k-1} \prod_{j'=1, \dots, m: j' \neq j} (\lambda_{j'} I - T)^n$$

to each side of (1.4.64) yields

$$a_j (\lambda_j I - T)^{k-1} \prod_{j'=1, \dots, m: j' \neq j} (\lambda_{j'} I - T)^n (v_j) = 0. \tag{1.4.65}$$

(Here we used Lemma 1.4.55.) For each $j' = 1, \dots, m$, $j' \neq j$, we write $\lambda_{j'} I - T = (\lambda_{j'} - \lambda_j) I + (\lambda_j I - T)$. Expanding (1.4.65) using the Multinomial Theorem gives

$$a_j \left\{ (\lambda_j I - T)^{k-1} \prod_{j'=1, \dots, m: j' \neq j} (\lambda_{j'} - \lambda_j)^n + \text{terms divisible by } (\lambda_j I - T)^k \right\} (v_j) = 0.$$

Hence

$$a_j \underbrace{\prod_{j'=1, \dots, m: j' \neq j} (\lambda_{j'} - \lambda_j)^n}_{\neq 0} \underbrace{(\lambda_j I - T)^{k-1}(v_j)}_{\neq 0} = 0$$

so $a_j = 0$. □

Lemma 1.4.66. *The collection of all generalized eigenvectors spans V .*

Proof. We prove this by induction on $n = \dim V$. The case $n = 1$ is obvious.

Given V and an eigenvalue λ of $T \in L(V, V)$, we begin by establishing the decomposition

$$V = \ker(\lambda I - T)^n \oplus \text{range}(\lambda I - T)^n. \quad (1.4.67)$$

Let $V_1 = \ker(\lambda I - T)^n$ and $V_2 = \text{range}(\lambda I - T)^n$ and suppose that $v \in V_1 \oplus V_2$. Then $(\lambda I - T)^n v = 0$ and $(\lambda I - T)^n u = v$ for some u . But then $(\lambda I - T)^{2n} u = 0$ so u is a generalized eigenvector for T and so (by Lemma 1.4.55) $(\lambda I - T)^n u = 0$. Thus $v = 0$. The Rank-Nullity theorem gives the equality of dimensions $n = \dim V = \dim V_1 + \dim V_2$ which, together with the fact that $V_1 \cap V_2 = \{0\}$, shows (1.4.67).

Since λ is an eigenvalue of T , $V_1 \neq 0$ so $\dim V_2 < n$. Thus the inductive hypothesis applies to V_2 and $T|_{V_2} \in L(V_2, V_2)$. We conclude that V_2 is spanned by generalized eigenvectors of $T|_{V_2}$, hence also of T . Clearly, V_1 is also spanned by generalized eigenvectors of T . This finishes the proof. □

Proof of Theorem 1.4.57(1). The decomposition $V = N^1 \oplus \dots \oplus N^m$ follows from Lemmas 1.4.63 and 1.4.66. The fact that $T : N^j \rightarrow N^j$ is a consequence of the observation that T commutes with $(\lambda_j I - T)^n$ for each j . Finally, it is clear that $(\lambda_j I - T)^n|_{N^j} = 0$. This finishes the proof. □

Exercise 1.4.68. Give an alternate proof of the decomposition $V = N^1 \oplus \dots \oplus N^m$ using the division algorithm for polynomials:

If P and Q are polynomials in $\mathbb{F}[\lambda]$ with no common zeros, then there exist $A, B \in \mathbb{F}[\lambda]$ so that $AP + BQ \equiv 1$.

(Hint: Induct to get a similar statement for m -tuples of polynomials P_1, \dots, P_m with no common zeros. Then apply this to the polynomials $P_1(\lambda) = (\lambda - \lambda_1)^{d(\lambda_1)}, \dots, P_m(\lambda) = (\lambda - \lambda_m)^{d(\lambda_m)}$.)

Proof of Theorem 1.4.57(2). We will prove the identity (1.4.59). The other identity (1.4.58) will follow as a consequence of the Jordan canonical form in part (3) of the theorem.

Form the polynomial

$$Q(\lambda) = \prod_j (\lambda - \lambda_j)^{d(\lambda_j)}.$$

In order to show that $Q = m_T$, it suffices to prove that $Q(T) = 0$ and that Q divides any polynomial P with $P(T) = 0$.

Let $v \in V$ be arbitrary. By part (1), we may write $v = \sum_{j'} v_{j'}$ with $v_{j'} \in N^{j'}$. In the j' th term of the expansion

$$Q(T)(v) = \sum_{j'} \prod_j (T - \lambda_j I)^{d(\lambda_j)}(v_{j'})$$

we commute the operators $(T - \lambda_j I)^{d(\lambda_j)}$ so that the term with $j = j'$ acts on $v_{j'}$ first. Since that term annihilates $v_{j'}$ we find $Q(T)(v) = 0$. Thus $Q(T) = 0$ as desired.

Next, suppose that $P \in \mathbb{F}[\lambda]$ satisfies $P(T) = 0$. Fix an index j and write

$$P(\lambda) = c \prod_{z_k \neq \lambda_j} (\lambda - z_k)^{\delta_k} \cdot (\lambda - \lambda_j)^\delta$$

for some values $z_k, c \in \mathbb{F}$ and nonnegative exponents δ_k, δ . We may also assume that $c \neq 0$. Let $v \in N^j$ be arbitrary. Then $(T - \lambda_j I)^\delta(v) \in N^j$ and $P(T)(v) = 0$. By Exercise 1.4.61 the operator $c \prod_k (T - z_k I)^{\delta_k}$ is one-to-one on N^j , so $(T - \lambda_j I)^\delta(v) = 0$. Since v was arbitrary, $(T - \lambda_j I)^\delta$ annihilates all of N^j , so $\delta \geq d(\lambda_j)$. Repeating this for each index j proves that Q divides P . This finishes the proof. \square

To prove part (3) of the theorem, we need the following

Lemma 1.4.69. *Let $T \in L(V, V')$ and let $W \subset V$, $W' \subset V'$ be subspaces so that $W = T^{-1}(W')$ and $\ker(T) \subset W$. Then there exists a one-to-one map $\tilde{T} \in L(V/W, V'/W')$ so that $Q' \circ T = \tilde{T} \circ Q$, where $Q : V \rightarrow V/W$ and $Q' : V' \rightarrow V'/W'$ are the canonical quotient maps.*

The proof of this lemma is an easy extension of Homework #1, Problem 3(a), apart from the claim that \tilde{T} is one-to-one. Let us sketch the proof of that claim. The map \tilde{T} is defined by the identity $\tilde{T}(v + W) := T(v) + W'$. If

$$\tilde{T}(v_1 + W) = \tilde{T}(v_2 + W),$$

then $T(v_1) + W' = T(v_2) + W'$ so $T(v_1 - v_2) \in W'$ and $v_1 - v_2 \in \ker(T) + T^{-1}(W') = \ker(T) + W = W$. Thus $v_1 + W = v_2 + W$ and we have shown that \tilde{T} is injective.

Proof of Theorem 1.4.57(3). For the purposes of this proof, we fix j and write $d = d_j$, $\lambda = \lambda_j$, $N = N^j$, and so on.

We begin by choosing a basis $\{v_{i,d}^j\}_i$ for $N_d(\lambda)/N_{d-1}(\lambda)$. We now apply Lemma 1.4.69 with $V = N_d(\lambda)$, $W = V' = N_{d-1}(\lambda)$, $W' = N_{d-2}(\lambda)$ and with the transformation in the lemma (called T there) given by $T - \lambda I$. Check that all of the hypotheses are verified. The lemma ensures that the quotient map

$$\widetilde{T - \lambda I} : N_d(\lambda)/N_{d-1}(\lambda) \rightarrow N_{d-1}(\lambda)/N_{d-2}(\lambda)$$

is one-to-one. Thus the images $(T - \lambda I)(v_{i,d}^j + N_{d-1}(\lambda))$ are linearly independent in $N_{d-1}(\lambda)/N_{d-2}(\lambda)$; we write these in the form $v_{i,d-1}^j + N_{d-2}(\lambda)$ and complete them to a basis $\{v_{i,d-1}^j\}_i$ of $N_{d-1}(\lambda)/N_{d-2}(\lambda)$. We repeat the process inductively, constructing bases $\{v_{i,l}^j\}_i$ for $N_l(\lambda)/N_{l-1}(\lambda)$. (In the final case we assume $N_0(\lambda) = \{0\}$.) The result of all of this is a set of vectors $\{v_{i,l}^j\}$ which forms a basis for $N = N^j$. (See Homework #1, Problem 3(b) for a similar claim.) \square

Proof of Theorem 1.4.57(4). By part (3),

$$T(v_{il}^j) = \begin{cases} \lambda_j v_{il}^j + v_{i,l-1}^j & \text{if } 1 < l \leq p_i^j, \\ \lambda_j v_{il}^j & \text{if } l = 1 \end{cases}$$

for all $j = 1, \dots, m$, $i = 1, \dots, k_j$. Consequently, the matrix representation of T with respect to the basis B consists of a block diagonal matrix made up of Jordan blocks with diagonal entries λ_j , as asserted. \square

1.5 Exponentiation of matrices and calculus of vector- and matrix-valued functions

Throughout this section, all matrices are defined over the complex numbers.

Definition 1.5.70. The *Hilbert-Schmidt norm* on $n \times n$ matrices is

$$\|A\|_{HS} := \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2}, \quad A = (a_{ij}).$$

Exercise 1.5.71. Verify that $\|\cdot\|_{HS}$ is a norm, i.e., $\|0\|_{HS} = 0$, $\|cA\|_{HS} = |c| \|A\|_{HS}$, and $\|AB\|_{HS} \leq \|A\|_{HS} \|B\|_{HS}$ for all $n \times n$ matrices A, B and $c \in \mathbb{F}$.

Definition 1.5.72. Let A be an $n \times n$ matrix. The *matrix exponential* e^A is defined to be

$$e^A = \sum_{j=0}^{\infty} \frac{1}{j!} A^j.$$

Lemma 1.5.73. *The infinite sum in the definition of e^A converges.*

Proof. Let $E_N(A) := \sum_{j=0}^N \frac{1}{j!} A^j$ denote the N th partial sum. Then

$$E_M(A) - E_N(A) = \sum_{j=M+1}^N \frac{1}{j!} A^j$$

so

$$\|E_M(A) - E_N(A)\|_{HS} \leq \sum_{j=M+1}^N \frac{1}{j!} \|A\|_{HS}^j \rightarrow 0 \quad \text{as } M, N \rightarrow \infty.$$

The conclusion follows. □

Exercise 1.5.74. Show that $e^{A+B} = e^A e^B$ if A and B commute. Show that in general e^{A+B} and $e^A e^B$ can disagree.

Lemma 1.5.75. *Let A, S be $n \times n$ matrices with S invertible. For every power series $f(t) = \sum_j c_j t^j$ converging on \mathbb{C} we have $f(S^{-1}AS) = S^{-1}f(A)S$.*

Proof. By uniform convergence it suffices to check this for polynomials. By linearity it suffices to check this for monomials. Finally,

$$(S^{-1}AS)^j = \underbrace{S^{-1}AS S^{-1}A \cdots AS S^{-1}AS}_{j \text{ repetitions}} = S^{-1}A^j S.$$

□

How do we calculate e^A in practice? Let $B = S^{-1}AS$ be the Jordan canonical form for A . Thus B is a block diagonal matrix with blocks in Jordan form. In this case $e^A = S e^B S^{-1}$, so it suffices to describe e^B for matrices B in Jordan canonical form. Matrix exponentiation factors across block diagonal decomposition of matrices, so it suffices to describe e^J for a matrix J consisting of a single Jordan block.

Proposition 1.5.76. Let J be an $n \times n$ Jordan block:

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

Then

$$e^J = \begin{pmatrix} e^\lambda & e^\lambda & \frac{1}{2}e^\lambda & \cdots & \frac{1}{(n-1)!}e^\lambda \\ 0 & e^\lambda & e^\lambda & \cdots & \frac{1}{(n-2)!}e^\lambda \\ 0 & 0 & e^\lambda & \cdots & \frac{1}{(n-3)!}e^\lambda \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & e^\lambda \end{pmatrix}.$$

Proof. Let

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

so $J = \lambda I + A$. By Exercise 1.5.74, $e^J = e^{\lambda I} e^A = e^\lambda (e^A)$. Since A is nilpotent of step n ,

$$e^A = I + A + \frac{1}{2}A^2 + \cdots + \frac{1}{(n-1)!}A^{n-1} = \begin{pmatrix} 1 & 1 & \frac{1}{2} & \cdots & \frac{1}{(n-1)!} \\ 0 & 1 & 1 & \cdots & \frac{1}{(n-2)!} \\ 0 & 0 & 1 & \cdots & \frac{1}{(n-3)!} \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

This finishes the proof. \square

The theory of matrix exponentials can also be developed via systems of ODE's.

Theorem 1.5.77. Let A be an $n \times n$ matrix and let $j \in \{1, \dots, n\}$. The j th column of e^A is the solution $x(1)$ at time $t = 1$ to the ODE

$$x'(t) = Ax(t), \quad x(0) = e_j.$$

Proof. Expanding both sides as series shows that $x(t) = e^{tA}x_0$ is the solution to the initial value problem

$$x'(t) = Ax(t), \quad x(0) = x_0.$$

Choosing $x_0 = e_j$ and evaluating at $t = 1$ finishes the proof. \square

The function $t \mapsto e^{tA}$ is an example of a *matrix-valued function*. To finish this section, we describe a few results from the calculus of such functions.

We consider functions $v : [a, b] \rightarrow \mathbb{F}^n$ and $A : [a, b] \rightarrow \mathbb{F}^{n \times n}$ taking values in a finite-dimensional vector space or the space of $n \times n$ matrices over \mathbb{F} . Continuity and differentiability of such functions is defined in the usual way. (We use the usual Euclidean norm on \mathbb{F}^n , and the Hilbert-Schmidt norm on $\mathbb{F}^{n \times n}$.) Most of the standard rules of differentiation hold true, but there are some complications in the matrix-valued case due to the possibility of non-commutativity. The chain rule for differentiating matrix-valued maps need not hold in general. Indeed, if $A : [a, b] \rightarrow \mathbb{F}^{n \times n}$ is a differentiable matrix-valued map, then

$$\frac{d}{dt}A^j = \dot{A} \cdot A^{j-1} + A \dot{A} A^{j-2} + \cdots + A^{j-1} \cdot \dot{A},$$

where $\dot{A} = \frac{dA}{dt}$. In particular, if $A(t)$ and $\dot{A}(t)$ commute for some value of t , then $\frac{d}{dt}A^j = jA^{j-1}\dot{A}$ and

$$\frac{d}{dt}p(A)(t) = p'(A)(t)\dot{A}(t)$$

for every polynomial p . Even if $A(t)$ and $\dot{A}(t)$ do not commute, we still have

$$\frac{d}{dt} \operatorname{tr} p(A)(t) = \operatorname{tr}(p'(A)(t)\dot{A}(t))$$

since the trace function is commutative.

To conclude, we describe a remarkable connection between determinant and trace. Roughly speaking, trace is the derivative of determinant. Consider the function $t \mapsto \det A(t)$ for some matrix-valued function A satisfying $A(0) = I$. We may write $A(t) = (a_1(t), \dots, a_n(t))$ in terms of columns with $a_j(t) = e_j$, and view $t \mapsto \det(a_1(t), \dots, a_n(t))$ as a multilinear function valued on n -tuples of vectors. Then

$$\frac{d}{dt} \det(a_1, \dots, a_n) = \det(\dot{a}_1, a_2, \dots, a_n) + \cdots + \det(a_1, a_2, \dots, \dot{a}_n);$$

evaluating at $t = 0$ gives

$$\begin{aligned} \left. \frac{d}{dt} \det A(t) \right|_{t=0} &= \det(\dot{a}_1(0), e_2, \dots, e_n) + \cdots + \det(e_1, e_2, \dots, \dot{a}_n(0)) \\ &= \dot{a}_{11}(0) + \cdots + \dot{a}_{nn}(0) = \operatorname{tr} \dot{A}(0). \end{aligned}$$

More generally, if $Y(t)$ is any differentiable matrix-valued function and $Y(t_0)$ is invertible, then

$$\left. \frac{d}{dt} \log \det Y(t) \right|_{t=t_0} = \operatorname{tr} \left(Y^{-1}(t_0) \dot{Y}(t_0) \right).$$

To prove this, set $A(t) = Y(t_0)^{-1}Y(t_0 + t)$ and apply the preceding result.