

8. Linear Maps

In this chapter, we study the notion of a linear map of abstract vector spaces. This is the abstraction of the notion of a linear transformation on \mathbf{R}^n .

8.1. Basic definitions

Definition 8.1. *Let U, V be vector spaces. A linear map $L : U \rightarrow V$ (reads L from U to V) is a rule which assigns to each element u an element in V , such that L preserves addition and scaling. That is,*

$$L(u_1 + u_2) = L(u_1) + L(u_2), \quad L(cu) = cL(u).$$

Notation: we sometimes write $L : u \mapsto v$, if $L(u) = v$, and we call v the image of u under L .

The two defining properties of a linear map are called *linearity*. Where does a linear map $L : U \rightarrow V$ send the zero element O in U ? By linearity, we have

$$L(O) = L(0O) = 0L(O).$$

But $0L(O)$ is the zero element, which we also denote by O , in V . Thus

$$L(O) = O.$$

Example. For any vector space U , we have a linear map $Id_U : U \rightarrow U$ given by

$$Id_U : u \mapsto u.$$

This is called the *identity map*. For any other vector space V , we also have a linear map $\mathbf{O} : U \rightarrow V$ given by

$$\mathbf{O} : u \mapsto \mathbf{O}.$$

This is called the *zero map*.

Example. Let A be an $m \times n$ matrix, and define

$$L_A : \mathbf{R}^n \rightarrow \mathbf{R}^m, \quad L_A(X) = AX.$$

We have seen, in chapter 3, that this map is linear. We have also seen that every linear map $L : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is represented by a matrix A , ie. if $L = L_A$ for some $m \times n$ matrix A .

Example. Let A an $m \times n$ matrix, and define

$$L_A : M(n, k) \rightarrow M(m, k), \quad L_A(B) = AB.$$

This map is linear because

$$A(B_1 + B_2) = AB_1 + AB_2, \quad A(cB) = c(AB)$$

for any $B_1, B_2, B \in M(n, k)$ and any number c (chapter 3).

Exercise. Is $L : \mathbf{R}^n \rightarrow \mathbf{R}$ defined by $L(X) = X \cdot X$, linear?

Exercise. Fix a vector A in \mathbf{R}^n . Is $L : \mathbf{R}^n \rightarrow \mathbf{R}$ defined by $L(X) = A \cdot X$, linear?

Example. (With calculus) Let C^∞ be the vector space of smooth functions, and let $D(f) = f'$. Then $D : C^\infty \rightarrow C^\infty$ is linear. differentiation

Example. (With calculus) Continuing with the preceding example, let $V_n = \text{Span}\{1, t, \dots, t^n\}$ in C^∞ . Since $D(t^k) = kt^{k-1}$ for $k = 0, 1, 2, \dots$, it follows that $D(f)$ lands in V_{n-1} for any function f in V_n . Thus we can regard D as a linear map

$$D : V_n \rightarrow V_{n-1}.$$

Theorem 8.2. Let U, V be vector spaces, S be a basis of U . Let $f : S \rightarrow V$ be any map. Then there is a unique linear map $L : U \rightarrow V$ such that $L(u) = f(u)$ for every element u

in S .

Proof: If u_1, \dots, u_k are distinct elements in S , then we define

$$L\left(\sum_{i=1}^k a_i u_i\right) = \sum_{i=1}^k a_i f(u_i)$$

for any numbers a_i . Since every element w in U can be expressed uniquely in the form $\sum_{i=1}^k a_i u_i$, this defines a map $L : U \rightarrow V$. Moreover, $L(u) = l(u)$ for every u in S . So it remains to verify that L is linear.

Let w_1, w_2, w be elements in U , and c be a scalar. We can write

$$\begin{aligned} w_1 &= \sum_{i=1}^k a'_i u_i \\ w_2 &= \sum_{i=1}^k a''_i u_i \\ w &= \sum_{i=1}^k a_i u_i \end{aligned}$$

where u_1, \dots, u_k are distinct elements in S , and the a 's are numbers. Then

$$\begin{aligned} L(w_1 + w_2) &= L\left[\sum (a'_i + a''_i)u_i\right] \\ &= \sum (a'_i + a''_i)f(u_i) \\ &= \sum a'_i f(u_i) + \sum a''_i f(u_i) \\ &= L(w_1) + L(w_2) \\ L(cw) &= L\left[c \sum a_i u_i\right] \\ &= L\left[\sum (ca_i)u_i\right] \\ &= \sum (ca_i)f(u_i) \\ &= c \sum a_i f(u_i) \\ &= cL(w). \end{aligned}$$

Thus L is linear. \square

For the rest of this section, $L : U \rightarrow V$ will be a linear map.

Definition 8.3. *The set*

$$\text{Ker } L = \{u \in U \mid L(u) = O\}$$

is called the kernel of L . The set

$$\text{Im } L = \{L(u) | u \in U\}$$

is called the image of L . More generally, for any subspace $W \subset U$, we write

$$L(W) = \{L(w) | w \in W\}.$$

Lemma 8.4. *Ker L is linear subspace of U . Im L is a linear subspace of V .*

Proof: If $u_1, u_2 \in \text{Ker } L$, then

$$L(u_1 + u_2) = L(u_1) + L(u_2) = O + O = O.$$

Thus $u_1 + u_2 \in \text{Ker } L$. If c is any number and $u \in \text{Ker } L$, then

$$L(cu) = cL(u) = cO = O.$$

Thus $cu \in \text{Ker } L$. Thus $\text{Ker } L$ is closed under addition and scaling in U , hence is a linear subspace of U .

The proof for $\text{Im } L$ is similar and is left as an exercise. \square

Example. Let A an $m \times n$ matrix, and $L_A : \mathbf{R}^n \rightarrow \mathbf{R}^m$ with $L_A(X) = AX$. Then $\text{Ker } L_A = \text{Null}(A)$, and $\text{Im } L_A = \text{Row}(A^t)$. This shows that the notions of kernel and image are abstractions of the notions of null space and row space.

Example. (With calculus) Consider $D : C^\infty \rightarrow C^\infty$ with $D(f) = f'$. $\text{Ker } D$ consists of just constant functions. Any smooth function is the derivative of a smooth function (fundamental theorem of calculus). Thus $\text{Im } D = C^\infty$.

Example. (With calculus) As before, let $V_n = \text{Span}\{1, t, \dots, t^n\}$ in C^∞ , and consider $D : V_n \rightarrow V_{n-1}$. As before, $\text{Ker } D$ consists of the constant functions. Since $D(t^k) = kt^{k-1}$ for $k = 0, 1, 2, \dots$, it follows $1, t, \dots, t^{n-1}$ are all in $\text{Im } D$. Thus any of their linear combination is also in $\text{Im } D$. So we have $\text{Im } D = V_{n-1}$.

Example. Fix a nonzero vector A in \mathbf{R}^n , and define

$$L : \mathbf{R}^n \rightarrow \mathbf{R}, \quad L : X \mapsto A \cdot X.$$

Then $\text{Ker } L = V^\perp$ where V is the line spanned by A , and $\text{Im } L = \mathbf{R}$.

Example. Let $L : U \rightarrow V$ be a linear map, and W be a linear subspace of U . We define a new map $L|_W : W \rightarrow V$ as follows:

$$L|_W(w) = L(w).$$

This map is linear. $L|_W$ is called the *restriction* of L to W .

8.2. A dimension relation

Throughout this section, $L : U \rightarrow V$ will be a linear map of finite dimensional vector spaces.

Lemma 8.5. *Suppose that $\text{Ker } L = \{O\}$ and that $\{u_1, \dots, u_k\}$ is a linearly independent set in U . Then $L(u_1), \dots, L(u_k)$ form a linearly independent set.*

Proof: Let

$$\sum_{i=1}^k x_i L(u_i) = O.$$

By linearity of L , this reads

$$L\left(\sum_{i=1}^k x_i u_i\right) = O.$$

Since $\text{Ker } L = \{O\}$, this implies that

$$\sum_{i=1}^k x_i u_i = O.$$

Since $\{u_1, \dots, u_k\}$ is linearly independent, it follows that the x 's are all zero. Thus $L(u_1), \dots, L(u_k)$ form a linearly independent set. \square

Theorem 8.6. (*Rank-nullity relation*) $\dim U = \dim (\text{Ker } L) + \dim (\text{Im } L)$.

Proof:

If $\text{Im } L$ is the zero space, then $\text{Ker } L = U$, and the theorem holds trivially. Suppose $\text{Im } L$ is not the zero space, and let $\{v_1, \dots, v_s\}$ be a basis of $\text{Im } L$. Let $L(u_i) = v_i$.

Let $\{w_1, \dots, w_r\}$ be a basis of $\text{Ker } L \subset U$. We will show that the elements $u_1, \dots, u_s, w_1, \dots, w_r$ form a linearly independent set that spans U .

• Linear independence: Let

$$x_1u_1 + \dots + x_su_s + y_1w_1 + \dots + y_rw_r = O.$$

Applying L to this, we get

$$x_1v_1 + \dots + x_sv_s = O.$$

Since $\{v_1, \dots, v_s\}$ is linearly independent, the x 's are all zero, and so

$$y_1w_1 + \dots + y_rw_r = O.$$

Since $\{w_1, \dots, w_r\}$ is linearly independent, the y 's are all zero. This shows that $u_1, \dots, u_s, w_1, \dots, w_r$ form a linearly independent set.

• Spanning property: Let $u \in U$. Since $\{v_1, \dots, v_s\}$ spans $\text{Im } L$, we have

$$L(u) = a_1v_1 + \dots + a_sv_s$$

for some a_i . This gives $L(u) = a_1L(u_1) + \dots + a_sL(u_s)$. By linearity of L , this yields

$$L(u - a_1u_1 - \dots - a_su_s) = 0,$$

which means that $u - a_1u_1 - \dots - a_su_s \in \text{Ker } L$. So

$$u - a_1u_1 - \dots - a_su_s = b_1w_1 + \dots + b_rw_r$$

because $\{w_1, \dots, w_r\}$ spans $\text{Ker } L$. This shows that

$$u = a_1u_1 + \dots + a_su_s + b_1w_1 + \dots + b_rw_r.$$

This completes the proof. \square

Example. Let A be an $m \times n$ matrix, and consider $L_A : \mathbf{R}^n \rightarrow \mathbf{R}^m$, $L_A(X) = AX$. Since

$$\text{Ker } L_A = \text{Null}(A), \quad \text{Im } L_A = \text{Row}(A^t), \quad \text{rank } A = \dim \text{Row}(A^t),$$

The preceding theorem yields

$$n = \text{rank } A + \dim \text{Null}(A).$$

We have proven this in chapter 4 by using an orthogonal basis. Note that the preceding theorem does not involve using any inner product.

Example. (With calculus) As before, let $V_n = \text{Span}\{1, t, \dots, t^n\}$ in C^∞ , and consider $D : V_n \rightarrow V_{n-1}$. Note that $\dim V_n = n + 1$. We have seen that $\text{Ker } D$ consists of the constant functions, and that $\text{Im } D = V_{n-1}$. So $\dim(\text{Ker } D) = 1$ and $\dim(\text{Im } D) = n$. Thus, in this case,

$$\dim V_n = \dim(\text{Ker } D) + \dim(\text{Im } D)$$

as expected.

Corollary 8.7. *Let U, W be finite dimensional linear subspaces of V . Then*

$$\dim(U) + \dim(W) = \dim(U + W) + \dim(U \cap W).$$

Proof: Define a linear map $L : U \oplus W \rightarrow V$, $(u, w) \mapsto u - w$. Then

$$\text{Ker } L = \{(u, u) \mid u \in U, u \in W\}, \quad \text{Im } L = U + W.$$

We have seen that $\dim(U \oplus W) = \dim(U) + \dim(W)$. By the dimension relation, it suffices to show that $\dim(\text{Ker } L) = \dim(U \cap W)$. For this, let $\{u_1, \dots, u_s\}$ be a basis of $U \cap W$, so that $\dim(U \cap W) = s$. We will show that the elements $(u_1, u_1), \dots, (u_s, u_s)$ form a basis of $\text{Ker } L$, so that $\dim(\text{Ker } L) = s$.

• Linear independence: consider

$$\sum x_i(u_i, u_i) = (O, O).$$

This means that

$$\sum x_i u_i = O.$$

Since $\{u_1, \dots, u_s\}$ is linearly independent, the x 's are all zero. So $(u_1, u_1), \dots, (u_s, u_s)$ form a linearly independent set.

• Spanning property: recall that an element of $\text{Ker } L$ is of the form (u, u) with $u \in U \cap W$. Since $\{u_1, \dots, u_s\}$ spans $U \cap W$, we have

$$u = \sum a_i u_i$$

for some scalars a_i . Thus

$$(u, u) = \left(\sum a_i u_i, \sum a_i u_i \right) = \sum a_i (u_i, u_i).$$

So $\{(u_1, u_1), \dots, (u_s, u_s)\}$ spans $\text{Ker } L$. \square

Definition 8.8. Let $L : U \rightarrow V$ be a linear map. We call L injective if $\text{Ker } L = 0$. We call L surjective if $\text{Im } L = V$. We call L bijective if L is both injective and surjective.

Corollary 8.9. Let $L : U \rightarrow V$ be a linear map of finite dimensional vector spaces.

(a) If L is injective then $\dim(U) \leq \dim(V)$.

(b) If L is surjective then $\dim(U) \geq \dim(V)$.

(c) If L is bijective then $\dim(U) = \dim(V)$.

Proof: All three follows from the dimension relation immediately. \square

Corollary 8.10. Let $L : U \rightarrow V$ be a linear map of finite dimensional vector spaces with $\dim(U) = \dim(V)$.

(a) If L is injective then L is bijective.

(b) If L is surjective then L is bijective.

Proof: (a) For L injective ie. $\text{Ker } L = 0$, the dimension relation says that $\dim(U) = \dim(\text{Im } L)$. So our assumption gives $\dim(\text{Im } L) = \dim(V)$, ie. $\text{Im } L$ is a subspace of V of the same dimension as V . A corollary of the Dimension Theorem says that $V = \text{Im } L$.

(b) For L surjective ie. $\text{Im } L = V$, our assumption says that $\dim(U) = \dim(\text{Im } L)$. So the dimension relation gives $\dim(\text{Ker } L) = 0$. Hence L is injective. \square

Example. *Warning.* The preceding corollary fails when U, V are infinite dimensional. Consider for example $D : C^\infty \rightarrow C^\infty$ as before. We know that C^∞ is infinite dimensional. Since every smooth function is the derivative of a smooth function, D is surjective. But we know that $D(1) = 0$, and so D is not injective.

Exercise. Let $M : C^\infty \rightarrow C^\infty$ be the map $M : f \mapsto tf$, ie. multiplication by the function t . Show that M is injective, but not surjective.

Theorem 8.11. *Let $L : U \rightarrow V$ be an injective linear map. If u_1, \dots, u_k form a linearly independent set in U , then $L(u_1), \dots, L(u_k)$ form a linearly independent set in V .*

Proof: Consider the linear relation

$$\sum_i x_i L(u_i) = O.$$

By linearity, we can rewrite this as

$$L\left(\sum_i x_i u_i\right) = O.$$

By injectivity, it follows that $\sum_i x_i u_i = O$. Since u_1, \dots, u_k form a linearly independent set, it follows that x_1, \dots, x_k are all zero. \square

Corollary 8.12. *Let $L : U \rightarrow V$ be an injective linear map. If S is a basis of U , then the set $L(S) = \{L(u) | u \in S\}$ is a basis of $Im L$.*

Proof: By the preceding theorem the set $L(S)$ is linearly independent. Since $L(S)$ also spans $Im L$, it follows that $L(S)$ is a basis of $Im L$. \square

Corollary 8.13. *Let $L : U \rightarrow V$ be a bijective linear map of finite dimensional vector spaces. If $\{u_1, \dots, u_n\}$ is a basis of U , then $\{L(u_1), \dots, L(u_n)\}$ is a basis of V .*

Proof: Since L is injective, $\{L(u_1), \dots, L(u_n)\}$ is a basis of $Im L$ by the preceding lemma. Since L is surjective, $Im L = V$. \square

Exercise. Let $\{u_1, \dots, u_n\}, \{v_1, \dots, v_n\}$ be respective bases of two vector spaces U, V . Show that there is a unique bijective linear map $L : U \rightarrow V$ such that

$$L(u_i) = v_i, \quad i = 1, \dots, n.$$

Definition 8.14. Given a linear map $L : U \rightarrow V$, we define its rank to be $\dim(\text{Im } L)$.

Thus

$$\text{rank}(L) = \dim(U) - \dim(\text{Ker } L).$$

Let $\{u_1, \dots, u_r\}$ and $\{v_1, \dots, v_s\}$ be bases of U and V respectively. For any $u \in U$, $L(u)$ is a linear combination of $\{v_1, \dots, v_s\}$ in a unique way. In particular,

$$L(u_i) = a_{i1}v_1 + \dots + a_{is}v_s = \sum_j a_{ij}v_j$$

where (a_{i1}, \dots, a_{is}) are the coordinates of $L(u_i)$ relative to $\{v_1, \dots, v_s\}$. We call the $r \times s$ matrix $A_L = (a_{ij})$ the *matrix of L* relative to the bases $\{u_1, \dots, u_r\}$ and $\{v_1, \dots, v_s\}$.

Theorem 8.15. $\text{rank}(A_L) = \text{rank}(L)$.

We will not detail the proof here. The proof is by setting up a bijective linear map

$$\text{Ker } L \rightarrow \text{Null}(A_L^t).$$

This gives $\dim(\text{Ker } L) = \dim(\text{Null } A_L^t)$. Then the theorem follows from the dimension relation. We will return to the study of the correspondence $L \mapsto A_L$ later.

• **Warning.** Let A be an $m \times n$ matrix, and consider the linear map $L_A : \mathbf{R}^n \rightarrow \mathbf{R}^m$, $X \mapsto AX$. Let $B = (b_{ij})$ be the matrix of L_A relative to the standard bases of $\mathbf{R}^n, \mathbf{R}^m$, which we denote by $\{e_1, \dots, e_n\}$, $\{E_1, \dots, E_m\}$ respectively. (Note that B is an $n \times m$ matrix.) Then

$$Ae_i = L_A(e_i) = \sum_j b_{ij}E_j.$$

The right hand side is the i th row of B , while Ae_i is the i th column of A . Therefore

$$B = A^t.$$

Thus the matrix of L_A relative to the standard bases is A^t , and not A .

8.3. Composition

We define composition by mimicking matrix multiplication. Composition is an operations which takes two linear maps (with appropriate domains and ranges) as input and yields another linear map as output.

Definition 8.16. *Let $L : U \rightarrow V$ and $M : V \rightarrow W$ be linear maps. The composition of L and M is the map $M \circ L : U \rightarrow W$ defined by $(M \circ L)(u) = M[L(u)]$.*

Exercise. Show that the map $M \circ L$ in the preceding definition is linear.

Exercise. Prove that if $L : U \rightarrow V$, $L' : U \rightarrow V$, $M : V \rightarrow W$, $M' : V \rightarrow W$ are linear maps and c is any scalar, then

$$\begin{aligned}(L + L') \circ M &= L \circ M + L' \circ M \\ (cL) \circ M &= c(L \circ M) \\ L \circ (M + M') &= L \circ M + L \circ M' \\ L \circ (cM) &= c(L \circ M).\end{aligned}$$

Thus we say that composition is *bilinear*.

Exercise. Prove that if $N : U \rightarrow V$, $M : V \rightarrow W$, $L : W \rightarrow Z$ are linear maps, then

$$(L \circ M) \circ N = L \circ (M \circ N).$$

Thus we say that composition is *associative*.

We now define inverse by mimicking the inverse of a matrix.

Definition 8.17. *We say that a linear map $L : U \rightarrow V$ is invertible if there is a linear map $M : V \rightarrow U$ such that $L \circ M = id_V$ and $M \circ L = id_U$. In this case,*

- *we call M an inverse of L ;*
- *we also call an invertible linear map, a linear isomorphism;*

- we say that U and V are isomorphic.

Lemma 8.18. *Let $L : U \rightarrow V$ be invertible linear map. Then L is bijective.*

Proof: Let M be an inverse of L .

- Injectivity: Let $L(u) = 0$. Applying M to this, we get $u = O$. Thus $\text{Ker } L = \{O\}$, and hence L is injective.

- Surjectivity: Let $v \in V$. Then $L[M(v)] = v$. Thus $\text{Im } L = V$, and hence L is surjective.

□

Lemma 8.19. *If L is invertible, then L has a unique inverse.*

Proof: Let M, N be inverses of L . Applying L to $M(v) - N(v)$, we get O . By injectivity of L , $M(v) - N(v) = O$. Thus $M = N$. □

Lemma 8.20. *Let $L : U \rightarrow V$ be bijective linear map. Then L is invertible.*

Proof: Given $v \in V$ there is a $u \in U$ such that $L(u) = v$ by surjectivity. If $u, u' \in U$ have the same image, ie. $L(u) = L(u')$, then $L(u - u') = 0$, ie. $u - u' = 0$ by injectivity. So given $v \in V$ there is exactly one $u \in U$ with $L(u) = v$.

Define a map $M : V \rightarrow U$ by $M(v) = u$ where $L(u) = v$. By definition $M \circ L(u) = u$ for all u , ie. $M \circ L = id_U$. Also $L \circ M(v) = v$ for all V , ie. $L \circ M = id_V$. It remains to show that M is linear.

- Addition: Applying L to $M(v_1 + v_2) - M(v_1) - M(v_2)$, we get O . Since L is injective, it follows that $M(v_1 + v_2) - M(v_1) - M(v_2) = O$.

- Scaling: Applying L to $M(cv) - cM(v)$, we also get O . □

Lemma 8.21. *If U, V are finite dimensional vector spaces with $\dim U = \dim V$, then there is a bijective linear map $L : U \rightarrow V$.*

Proof: Let $n = \dim U = \dim V$ and $\{u_1, \dots, u_n\}, \{v_1, \dots, v_n\}$ be bases of U, V respectively. Define a map $f : \{u_1, \dots, u_n\} \rightarrow V, f(u_i) = v_i$ for all i . By Theorem 8.2, there is a linear map $L : U \rightarrow V$ such that $L(u_i) = v_i$. It remains to show that L is bijective. Since both U, V are finite dimensional, it suffices to show that L is surjective.

Let v be an element of V . Then v can be expressed as $v = \sum a_i v_i$. By linearity of L , we have

$$L\left(\sum a_i u_i\right) = \sum a_i L(u_i) = \sum a_i v_i = v.$$

Thus v lies in $\text{Im } L$. We conclude that $\text{Im } L = V$. \square

Example. Let $U = \text{Span}\{1, t, \dots, t^n\}$ in C^∞ . Let E_1, \dots, E_{n+1} be the standard unit vectors in \mathbf{R}^{n+1} . We have a bijective linear map $L : U \rightarrow \mathbf{R}^{n+1}$ such that $L(t^i) = E_{i+1}$ for $i = 0, 1, \dots, n$.

In summary, we have shown the following. For finite dimensional vector spaces U, V , the following statements are equivalent:

- (a) There is an invertible linear map $L : U \rightarrow V$.
- (b) There is a bijective linear map $L : U \rightarrow V$.
- (c) $\dim(U) = \dim(V)$.

Lemma 8.22. *If U, V are finite dimensional vector spaces of the same dimension and $L : U \rightarrow V, M : V \rightarrow U$ are linear maps with $L \circ M = \text{id}_V$, then L is invertible with inverse M .*

Proof: For any element v in V ,

$$L[M(v)] = v.$$

Thus L is surjective. By the dimension relation,

$$\dim(\text{Ker } L) + \dim(V) = \dim(U).$$

That U and V having the same dimension implies that $\text{Ker } L = \{O\}$. Thus L is injective. This means that L is invertible. By the associativity property of composition, we have

$$L^{-1} \circ (L \circ M) = (L^{-1} \circ L) \circ M = id_U \circ M = M.$$

But since $L \circ M = id_V$, we also have $L^{-1} \circ (L \circ M) = L^{-1}$. It follows that $M = L^{-1}$. \square

Exercise. Give an example to show that the statement without the assumption that $\dim(U) = \dim(V)$ in the lemma is false.

Exercise. Prove that if U, V are finite dimensional vector spaces of the same dimension and $L : U \rightarrow V$, $M : V \rightarrow U$ are linear maps with $M \circ L = id_U$, then L is invertible with inverse M .

8.4. Linear equations

Terminology: Let $L : U \rightarrow V$ be a linear map. An equation of the form

$$L(X) = v_0$$

is called a linear equation in the variable X . When the given v_0 is $\vec{0}$, we call that a homogeneous equation. We call a vector $u \in U$ a solution if $L(u) = v_0$. Note that we do not require that U, V be finite dimensional. For $U = \mathbf{R}^n$, $V = \mathbf{R}^n$, and $L = L_A$, $v_0 = B$, the linear equation $L_A(X) = B$ reads $AX = B$. This is what we used to call a linear system.

Theorem 8.23. *Suppose $L(u_0) = v_0$. Then the solution set of the equation $L(X) = v_0$ is*

$$\{w + u_0 | w \in \text{Ker } L\}.$$

Proof: Call this set S . A vector in S is of the form $w + u_0$ with $w \in \text{Ker } L$. So $L(w + u_0) = L(w) + L(u_0) = v_0$, hence $w + u_0$ is a solution. Conversely if u is a solution, then $L(u) = v_0$. Thus $L(u - u_0) = O$, and hence $w = u - u_0 \in \text{Ker } L$. It follows that $u = w + u_0$. \square

Example. (With calculus) *Differential equations.* Consider the linear map $D+cI : C^\infty \rightarrow C^\infty$ where $D(f) = f'$ as before, I is the identity map and c is a number. We want to solve the equation

$$(D + cI)f = O.$$

This is an example of a differential equation. Since $D(e^{-ct}) = -ce^{-ct}$, it follows that e^{-ct} is a solution. Let f be any solution, and let $g = fe^{ct}$. Then $f = ge^{-ct}$, and so $(D + cI)(ge^{-ct}) = O$. Explicitly, this reads

$$g'e^{-ct} = O$$

which shows that $g' = O$. Thus g is a constant function, and hence f is a scalar multiple of e^{-ct} . It follows that the solution space $\text{Ker}(D + cI)$ is spanned by the function e^{-ct} .

Let's solve the inhomogeneous equation

$$(D + cI)f = 1.$$

If $c \neq 0$, then $(D + cI)\frac{1}{c} = 1$. By the preceding theorem, the solution set of our equation in this case consists of functions of the form

$$\text{const. } e^{-ct} + \frac{1}{c}.$$

If $c = 0$, then $(D + cI)t = 1$. The solution set in this case consists of functions of the form

$$\text{const. } e^{-ct} + t.$$

8.5. The Hom space

Let U, V be vector spaces. Let L, M be two linear maps from U to V , and c be a scalar. We define the sum $L + M$ and the scalar multiple cL as the following maps from U to V :

$$L + M : U \rightarrow V, \quad (L + M)(u) = L(u) + M(u)$$

$$cL : U \rightarrow V, \quad (cL)(u) = cL(u).$$

- $L + M$ and cL are linear.

Proof: Let u_1, u_2, u be elements of U , and b be scalars. We chase through a chain of definitions as follows:

$$\begin{aligned}
 (L + M)(u_1 + u_2) &= L(u_1 + u_2) + M(u_1 + u_2) \\
 &= (L(u_1) + L(u_2)) + (M(u_1) + M(u_2)) \\
 &= (L(u_1) + M(u_1)) + (L(u_2) + M(u_2)) \\
 &= (L + M)(u_1) + (L + M)(u_2) \\
 (L + M)(bu) &= L(bu) + M(bu) \\
 &= bL(u) + bM(u) \\
 &= b[L(u) + M(u)] \\
 &= b[(L + M)(u)].
 \end{aligned}$$

This proves that $L + M$ is a linear map. Checking the linearity of cL is similar. \square

Let $\text{Hom}(U, V)$ denote the set of all linear maps from U to V . This set contains the zero map $\mathbf{0}$. We have just defined addition and scaling on this set.

- $\text{Hom}(U, V)$ is a vector space with addition and scaling defined above.

Proof: Verifying each of the axioms V1-V8 needs no more than chasing through definitions in a straightforward manner. We illustrate this for V1 and V5, and leave the rest as exercise.

Let L, M, N be three linear maps from U to V . To prove V1, we must show that the linear map $(L + M) + N$ is equal to the linear map $L + (M + N)$. Two linear maps being equal means that they both do the same thing to any given element u . Thus we are to show that

$$[(L + M) + N](u) = [L + (M + N)](u)$$

for any element u in U . Let's start from the left hand side and derive the right hand side:

$$\begin{aligned}
 [(L + M) + N](u) &= (L + M)(u) + N(u) \\
 &= (L(u) + M(u)) + N(u) \\
 &= L(u) + (M(u) + N(u)) \\
 &= L(u) + (M + N)(u) \\
 &= [L + (M + N)](u).
 \end{aligned}$$

This proves V1.

Let a be a scalar and L, M be two linear maps from U to V . We want to show that

$$[a(L + M)](u) = (aL + aM)(u)$$

for any element u in U . Again, let's start from the left:

$$\begin{aligned} [a(L + M)](u) &= a[(L + M)(u)] \\ &= a[L(u) + M(u)] \\ &= aL(u) + aM(u) \\ &= (aL)(u) + (aM)(u) \\ &= (aL + aM)(u). \end{aligned}$$

This proves V5. \square

Example. Consider the special case $V = \mathbf{R}$. The vector space $\text{Hom}(U, \mathbf{R})$ is often denoted by U^* , and is called the *linear dual* of U . An element of $U^* = \text{Hom}(U, \mathbf{R})$ is called a *linear form* or a 1-form on U .

Example. Let V be a 1 dimensional vector space, and $L : V \rightarrow V$ be a linear map. Then V consists of all scalar multiples cv_0 of some fixed element v_0 . Since $L(v_0)$ is an element of V , we have

$$L(v_0) = c_0v_0$$

for some number c_0 . It follows that

$$L(cv_0) = cL(v_0) = cc_0v_0 = (c_0Id_V)(cv_0)$$

for any number c . We conclude that $L = c_0Id_V$. This shows that *any linear map on a one dimensional vector space V is a scalar multiple of the identity map Id_V* . In particular,

$$\dim \text{Hom}(V, V) = 1.$$

Let U, V be finite dimensional vector spaces. Fix bases $\{u_1, \dots, u_r\}$ and $\{v_1, \dots, v_s\}$ of U and V respectively. For every element L in $\text{Hom}(U, V)$, let $A_L = (a_{ij})$ be the matrix of L relative to those bases. Thus

$$L(u_i) = \sum_{j=1}^s a_{ij}v_j.$$

Define a map

$$F : \text{Hom}(U, V) \rightarrow M(r, s), \quad F : L \mapsto A_L.$$

Note that this map depends on the choice of our bases. We call this *the matrix representation map* relative to the bases.

Theorem 8.24. *The matrix representation map F relative to any given basis is a bijective linear map.*

Proof:

• **Linearity:** Let L_1, L_2, L be elements of $\text{Hom}(U, V)$, and c be a scalar. Let $A_{L_1} = (a'_{ij})$, $A_{L_2} = (a''_{ij})$, $A_L = (a_{ij})$.

$$\begin{aligned} (L_1 + L_2)(u_i) &= L_1(u_i) + L_2(u_i) \\ &= \sum_{j=1}^s a'_{ij} v_j + \sum_{j=1}^s a''_{ij} v_j \\ &= \sum_{j=1}^s (a'_{ij} + a''_{ij}) v_j. \end{aligned}$$

This says that $a'_{ij} + a''_{ij}$ is the (ij) entry of the matrix of $A_{L_1+L_2}$. On the other hand, $a'_{ij} + a''_{ij}$ is also the (ij) entry of the matrix $A_{L_1} + A_{L_2}$. So we conclude that

$$A_{L_1+L_2} = A_{L_1} + A_{L_2},$$

or

$$F(L_1 + L_2) = F(L_1) + F(L_2).$$

Similarly

$$\begin{aligned} (cL)(u_i) &= cL(u_i) \\ &= c \sum_{j=1}^s a_{ij} v_j \\ &= \sum_{j=1}^s (ca_{ij}) v_j. \end{aligned}$$

This says that ca_{ij} is the (ij) entry of the matrix A_{cL} . On the other hand, ca_{ij} is also the (ij) entry of the matrix cA_L . So we conclude that

$$A_{cL} = cA_L$$

or

$$F(cL) = cF(L).$$

- Injectivity: Suppose that $F(L) = A_L$ is the zero matrix. Then

$$L(u_i) = O$$

for all i . By linearity of L , it maps any linear combination of $\{u_1, \dots, u_r\}$, and hence any element u in U , to O . Thus L is the zero map from U to V . Hence F is injective.

- Surjectivity: Let $B = (b_{ij})$ be any $r \times s$ matrix. Define the assignment $f : \{u_1, \dots, u_r\} \rightarrow V$, $f : u_i \mapsto \sum_{j=1}^s b_{ij}v_j$. By Theorem 8.2, there is a linear map $L : U \rightarrow V$ such that

$$L(u_i) = f(u_i) = \sum_{j=1}^s b_{ij}v_j.$$

This shows that $A_L = B$, ie. $F(L) = B$. Thus F is surjective.

This completes our proof. \square

Corollary 8.25. *If U and V are finite dimensional vector spaces, then $\dim \text{Hom}(U, V) = (\dim U)(\dim V)$.*

Proof: By the preceding theorem,

$$\dim \text{Hom}(U, V) = \dim M(r, s) = rs,$$

where $r = \dim U$ and $s = \dim V$. \square

Corollary 8.26. *If U is finite dimensional, then $\dim U^* = \dim U$.*

Proof: Since $U^* = \text{Hom}(U, \mathbf{R})$, it follows from the preceding corollary that

$$\dim U^* = (\dim U)(\dim \mathbf{R}) = \dim U. \quad \square$$

The following theorem says that the matrix representation map relative to any given bases has another important property: that it relates composition of linear maps with multiplication of their matrices.

Theorem 8.27. *Let U, V, W be finite dimensional vector spaces with dimension r, s, t respectively. Fix three respective bases, and let*

$$F : \text{Hom}(U, V) \rightarrow M(r, s), \quad G : \text{Hom}(V, W) \rightarrow M(s, t), \quad H : \text{Hom}(U, W) \rightarrow M(r, t)$$

be the respective matrix representation maps relative to those bases. Then for any linear maps $L : U \rightarrow V$, $M : V \rightarrow W$,

$$H(M \circ L) = F(L)G(M).$$

Proof: Let $\{u_1, \dots, u_r\}$, $\{v_1, \dots, v_s\}$, $\{w_1, \dots, w_t\}$ be bases we have chosen for the vector spaces U, V, W respectively. Let $N = M \circ L$. Let $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$ be the respective matrices of L, M, N relative to the given bases. We must show that

$$C = AB.$$

By definition of composition,

$$N(u_i) = M[L(u_i)] = \sum_j a_{ij} M(v_j) = \sum_j \sum_k a_{ij} b_{jk} w_k.$$

On the other hand, $N(u_i) = \sum_k c_{ik} w_k$. Comparing the coefficient of each w_k , we conclude that

$$c_{ik} = \sum_j a_{ij} b_{jk}$$

for all i, k . This gives the asserted matrix identity $C = AB$. \square

Corollary 8.28. *Let U be an n dimensional vector space. Fix a basis and let*

$$F : \text{Hom}(U, U) \rightarrow M(n, n), \quad L \mapsto A_L$$

be the matrix representation map relative to that basis. Then for any linear maps $L : U \rightarrow U$, $M : U \rightarrow U$,

$$A_{M \circ L} = A_L A_M.$$

Proof: This follows from specializing the preceding theorem to the case $U = V = W$. \square

Corollary 8.29. *Assume the same as in the preceding corollary. If $L : U \rightarrow U$ is invertible linear map, then A_L is an invertible matrix with inverse $A_{L^{-1}}$.*

Proof: This follows from specializing the preceding corollary to the case $M = L^{-1}$, and the fact that A_{Id} is the $n \times n$ identity matrix. \square

Exercise. Assume the same as in the preceding two corollaries. Prove that if L, M, N are any linear maps $U \rightarrow U$, then

$$A_{N \circ M \circ L} = A_L A_M A_N.$$

8.6. Induced maps

Let U, V, W be vector spaces. Let $f : V \rightarrow W$ be a given linear map. We define a map

$$H_f : \text{Hom}(U, V) \rightarrow \text{Hom}(U, W), \quad H_f : L \mapsto f \circ L.$$

This is called *the map induced by f* . Here is a diagram which is a good mnemonic device for this definition:

$$\begin{array}{ccc} & \xrightarrow{f \circ L} & \\ & \text{---} & \\ U & \xrightarrow{L} V \xrightarrow{f} & W. \end{array}$$

- H_f is a linear map.

Proof: Let L_1, L_2, L be elements of $\text{Hom}(U, V)$, and c be a scalar. We want to show that

$$H_f(L_1 + L_2) = H_f(L_1) + H_f(L_2)$$

$$H_f(cL) = cH_f(L).$$

Now $H_f(L_1 + L_2)$ and $H_f(L_1) + H_f(L_2)$ are linear maps from U to W . So to prove that they are equal means to show that

$$[H_f(L_1 + L_2)](u) = [H_f(L_1) + H_f(L_2)](u)$$

for any element u in U . Starting from the left, we have

$$\begin{aligned}
 [H_f(L_1 + L_2)](u) &= [f \circ (L_1 + L_2)](u) \\
 &= f[(L_1 + L_2)(u)] \\
 &= f[L_1(u) + L_2(u)] \\
 &= f[L_1(u)] + f[L_2(u)] \\
 &= (f \circ L_1)(u) + (f \circ L_2)(u) \\
 &= [H_f(L_1)](u) + [H_f(L_2)](u) \\
 &= [H_f(L_1) + H_f(L_2)](u).
 \end{aligned}$$

This proves that $H_f(L_1 + L_2)$ and $H_f(L_1) + H_f(L_2)$ are equal. Similarly,

$$\begin{aligned}
 [H_f(cL)](u) &= [f \circ (cL)](u) \\
 &= f[(cL)(u)] \\
 &= f[cL(u)] \\
 &= c f[L(u)] \\
 &= c (f \circ L)(u) \\
 &= [c(f \circ L)](u) \\
 &= [c H_f(L)](u).
 \end{aligned}$$

This proves that $H_f(cL) = cH_f(L)$, hence completes our proof that H_f is a linear map.

□

Let U, V, W be vector spaces. Let $f : U \rightarrow V$ be a given linear map. We define a map

$$H^f : \text{Hom}(V, W) \rightarrow \text{Hom}(U, W), \quad H^f : L \mapsto L \circ f.$$

The diagram for this definition is:

$$\begin{array}{ccc}
 & \xrightarrow{L \circ f} & \\
 U & \xrightarrow{f} V \xrightarrow{L} & W.
 \end{array}$$

Exercise. Prove that H^f is a linear map.

Exercise. Prove that if $f : V \rightarrow W$ is invertible, then $H_f : \text{Hom}(U, V) \rightarrow \text{Hom}(U, W)$ is also invertible with inverse $H_{f^{-1}}$.

Exercise. Prove that if $f : U \rightarrow V$ is invertible, then $H^f : \text{Hom}(V, W) \rightarrow \text{Hom}(U, W)$ is also invertible with inverse $H^{f^{-1}}$.

8.7. Tensor product spaces

Let S be any nonempty set. Let $F(S)$ be the set consisting of *all* functions

$$f : S \rightarrow \mathbf{R}$$

such that $f(s) = 0$ for all but finitely many $s \in S$. It is a vector space under the usual rules of function addition and scaling. It is called the *the free vector space on S* . The set $\{f_s | s \in S\}$, where $f_s(r) = 1$ for $r \neq s$ and $f_s(s) = 0$, is a basis of $F(S)$. It is called the *canonical basis* of $F(S)$. For any function $f : S \rightarrow \mathbf{R}$ can be uniquely written as a linear combination $\sum_s f(s)f_s$. It is convenient to denote a vector in $F(S)$ as a formal sum $\sum_s a_s s$, where it is understood that this represents the function $S \rightarrow \mathbf{R}$, $s \mapsto a_s$.

Exercise. Let $S = \{1, 2, \dots, n\}$. Show that there is a canonical linear isomorphism $F(S) \rightarrow \mathbf{R}^n$.

Now let V, W be vector spaces, and consider $F(V \times W)$. Let $R(V, W)$ be the linear subspace of $F(V \times W)$ generated by the vectors

$$(v_1 + cv_2, w) - (v_1, w) - c(v_2, w), \quad (v, w_1 + cw_2) - (v, w_1) - c(v, w_2)$$

($v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, $c \in \mathbf{R}$.)

Definition 8.30. *The tensor product space $V \otimes W$ is the quotient space $F(V \times W)/R(V, W)$.*

In particular, $V \otimes W$ is spanned by cosets of the form $v \otimes w := (v, w) + R(V, W)$.

Exercise. *The main point.* Verify that the expression $v \otimes w$ is linear in v and w .

Exercise. Let U be a vector space and $B : V \times W \rightarrow U$ is a bilinear map. Prove that there is a unique linear map $\tilde{B} : V \otimes W \rightarrow U$ such that $\tilde{B}(v \otimes w) = B(v, w)$ for all $(v, w) \in V \times W$. (Hint: First define a map $F(V \times W) \rightarrow U$ using the canonical basis of $F(V \times W)$, and then show that $R(V, W)$ lies in the kernel.)

8.8. Homework

1. Consider the linear map $D^2 = D \circ D : C^\infty \rightarrow C^\infty$. Describe $\text{Ker } D^2$ and $\text{Im } D^2$. Do the same for $D^k = D \circ D \circ \cdots \circ D$ (k times).

2. Define a map $T : M(m, n) \rightarrow M(n, m)$, $T(A) = A^t$. Show that T is linear.

3. Let $L : U \rightarrow \mathbf{R}$ be a linear map from a 4 dimensional vector space U to \mathbf{R} , what are the possible dimensions of $\text{Ker } L$? Give an example in each case.

4. Define the map

$$L : M(n, n) \rightarrow M(n, n), \quad L(A) = \frac{1}{2}(A + A^t).$$

(a) Show that L is linear.

(b) Show that $\text{Ker } L$ consists of all skew symmetric matrices.

(c) Show that $\text{Im } L$ consists of all symmetric matrices.

5. (a) Let A be a given $n \times n$ matrix. Let S be the vector space of $n \times n$ symmetric matrices. Define a map $L : S \rightarrow S$, by

$$L(X) = A^t X A.$$

Show that L is a linear map.

(b) Show that if A is an invertible matrix, then L is invertible.

6. Let $L : U \rightarrow V$ be a linear map of finite dimensional vector spaces.

(a) Show that $\dim U \geq \dim (\text{Im } L)$.

(b) Prove that if $\dim U > \dim V$, then $\text{Ker } L \neq 0$.

7. Let V be a vector space with an inner product $\langle \cdot, \cdot \rangle$. For every element v in V , we define a map $f_v : V \rightarrow \mathbf{R}$ by $f_v(u) = \langle v, u \rangle$.
- (a) Show that f_v is linear, hence it is an element of V^* .
- (b) Define a map $R : V \rightarrow V^*$ by $R(v) = f_v$. Show that R is linear.
- (c) Show that R is injective.
- (d) Show that if V is finite dimensional, then R is a linear isomorphism.

8. Let V be a finite dimensional vector space, and $\{u_1, \dots, u_n\}$ be a basis of V . For each i , we define a linear form u_i^* on V by

$$u_i^*(u_i) = 1, \quad u_i^*(u_j) = 0 \quad j \neq i.$$

Prove that u_1^*, \dots, u_n^* form a basis of V^* . It is called the basis dual to $\{u_1, \dots, u_n\}$.

9. Let $V = \mathbf{R}^n$ equipped with the dot product. Then we have a linear isomorphism $R : \mathbf{R}^n \rightarrow V^*$ as defined above. Let $\{E_1^*, \dots, E_n^*\}$ be the basis of V^* dual to the standard basis of \mathbf{R}^n . Find the images of the E_i^* under the linear map R^{-1} . Now do the same for \mathbf{R}^2 with the basis $\{(1, -1), (2, 1)\}$.
10. Let U, V be a vector space, and $L : U \rightarrow V$ be a linear map. Define a second map $L^* : V^* \rightarrow U^*$ as follows. Given an element f in V^* and u in U , let

$$[L^*(f)](u) = f[L(u)].$$

Show that L^* is linear. It is called the *dual map* of L .

11. Continuing with the preceding exercise, suppose in addition that U, V are finite dimensional. Let A_L be the matrix of L relative to some given bases of U and V respectively. Show that the matrix of $L^* : V^* \rightarrow U^*$ relative to the dual bases is the transpose matrix A_L^t .

12. (With calculus) *More on differential equations.* Consider the linear map $L = D^2 - 3D + 2I : C^\infty \rightarrow C^\infty$. We want to solve the equation

$$L(f) = 0.$$

- (a) Show that e^t and e^{2t} are solutions to the equation.
- (b) Let f be any solution, and let $g = fe^{-t}$. Show that $g'' - g' = 0$. Hence conclude that $g' = \text{const.}e^t$.
- (c) Show that f is a linear combination of e^t and e^{2t} . Hence conclude that $\text{Ker } L$ is spanned by these two functions.
- (d) Show that

$$L(1) = 2.$$

- (e) Describe all solutions to the equation

$$L(f) = 1.$$

13. Let V be any finite dimensional vector space and U_1, U_2 be two linear subspaces of V of the same dimension. Show that there is an invertible map $L : V \rightarrow V$ such that $L(U_1) = U_2$. (Hint: Start with bases of U_1, U_2 .)

14. * Consider the sequence of two linear maps

$$U \xrightarrow{f} V \xrightarrow{g} W.$$

We say that this sequence is exact if f is injective, g is surjective, and $\text{Im } f = \text{Ker } g$. Show that if U, V, W are finite dimensional and the sequence is exact, then there is a linear isomorphism

$$V \cong U \oplus W.$$

15. Let A be an $n \times n$ matrix having entries below and along the diagonal all zero. Regard $A : \mathbf{R}^n \rightarrow \mathbf{R}^n$ as a linear map as usual. (Cf. this problem with problem 6, Chapter 3.)

(a) For $1 \leq k \leq n$, let V_k be the linear subspace spanned by the standard unit vectors E_1, \dots, E_k in \mathbf{R}^n , and let V_0 be the zero subspace. Show that

$$AE_k \in V_{k-1}.$$

(b) Show that $AV_k \subset V_{k-1}$. Hence conclude that $A^i V_k \subset V_{k-i}$, for $1 \leq i \leq k \leq n$. Thus each power A^i has the same shape as A except that the band of zeros increases its width by one (or more) each time the power increases by one.

16. * Let V be a finite dimensional vector space and $A, B : V \rightarrow V$ be linear maps. Prove that

$$\dim \text{Ker}(A \circ B) \leq \dim \text{Ker}(A) + \dim \text{Ker}(B).$$

(Hint: Write $V = \text{Ker}(B) + U$ where U is a complementary subspace of $\text{Ker}(B)$ in V .)

17. * Let V be a finite dimensional inner product space and W a subspace of V .

(a) Let $T : V \rightarrow V$ be the linear map such that

$$T(w) = w, \quad w \in W; \quad T(u) = 0, \quad u \in W^\perp.$$

This is called the orthogonal *projection* of V onto W . Prove that T has the properties

(i) $T^2 = T$

(ii) $\text{Im } T = W$

(iii) $\|T(v)\| \leq \|v\|$.

Now suppose that $T : V \rightarrow V$ is a given linear map with properties (i)-(iii).

(b) Show that

$$V = \text{Im}(T) + \text{Im}(I - T)$$

and that the right side is a direct sum. Also show that

$$\text{Im}(T) = \text{Ker}(I - T), \quad \text{Im}(I - T) = \text{Ker}(T).$$

- (c) Show that $\text{Ker}(T)^\perp = \text{Ker}(I - T)$. (Hint: Assume that $w \in \text{Ker}(I - T)$, $u \in \text{Ker}(T)$ are not orthogonal. Show that this violates (iii) for $v = w + cu$, for an appropriate number c .)
- (d) Conclude that T is the orthogonal projection of V onto W .
18. Let V be a finite dimensional inner product space. Suppose U_1 and U_2 are subspaces of V with $\dim U_1 < \dim U_2$. Prove that U_2 contains a nonzero vector u such that $\langle u, w \rangle = 0$ for all vector w in U_1 . (Hint: Consider the map $U_2 \rightarrow U_1^*$, $u \mapsto \langle u, - \rangle$.)
19. * Let V and W be possibly infinite dimensional vector spaces. Show that there is a linear map $\alpha : V^* \otimes W \rightarrow \text{Hom}(V, W)$, such that $\lambda \otimes w \mapsto \lambda(-)w$, for all $\lambda \in V^* = \text{Hom}(V, \mathbf{R})$ and $w \in W$. Prove that α is injective, and that its image is the set consisting of $\rho : V \rightarrow W$ with $\dim \rho(V) < \infty$. Conclude that α is surjective iff either V or W is finite dimensional.
20. Use the preceding problem to prove that if V, W are finite dimensional, then $\dim V \otimes W = \dim V \cdot \dim W$.
21. Let $U \subset W$ and V be vector spaces and U, W finite dimensional. Suppose that $\omega \in W \otimes V \equiv \text{Hom}(W^*, V)$ defines an injective map $\tau_\omega : W^* \rightarrow V$, $\lambda \mapsto \lambda(\omega)$. Show that

$$\bar{\omega} = \omega + U \otimes V \subset W/U \otimes V \equiv W \otimes V/U \otimes V \equiv \text{Hom}((W/U)^*, V)$$

also defines an injective map $\tau_{\bar{\omega}} : (W/U)^* \rightarrow V$. Conclude that if $w_1, \dots, w_n \in W$ form a basis of W , $w'_1 + U, \dots, w'_m + U \in W/U$ form a basis of W/U , $v_1, \dots, v_n \in V$ are independent, and if

$$\sum_{i=1}^n w_i \otimes v_i \equiv \sum_{j=1}^m w'_j \otimes v'_j \pmod{U \otimes V}$$

then $v'_1, \dots, v'_m \in V$ are independent. (Hint: $\tau_{\bar{\omega}} = \tau_\omega \circ \pi^*$, $\pi^* : (W/U)^* \hookrightarrow W^*$.)

9. Determinants For Linear Maps

Throughout this chapter, V will be a finite dimensional vector space. We denote $\dim V$ by n .

9.1. p -forms

We denote by $V \times \cdots \times V$ (p times), the set of all p -tuples (v_1, \dots, v_p) of elements v_i in V .

Definition 9.1. A map $f : V \times \cdots \times V \rightarrow \mathbf{R}$ is called a p -form on V if it is linear in each slot, ie. for any elements v_1, \dots, v_p, v in V and any scalar c ,

$$\begin{aligned} f(v_1, \dots, v_i + v, \dots, v_p) &= f(v_1, \dots, v_i, \dots, v_p) + f(v_1, \dots, v, \dots, v_p) \\ f(v_1, \dots, cv_i, \dots, v_p) &= cf(v_1, \dots, v_i, \dots, v_p) \end{aligned}$$

for all i . By convention, a 0-form is a number. We denote by $\mathcal{T}^p(V)$ the set of p -forms on V . We say that f is symmetric if

$$f(v'_1, \dots, v'_p) = f(v_1, \dots, v_p)$$

(v'_1, \dots, v'_p) is obtained from (v_1, \dots, v_p) by swapping two of the v 's. We denote by $\mathcal{S}^p(V)$ the set of all symmetric p -forms on V . We say that f is skew-symmetric if

$$f(v_1, \dots, v_p) = 0$$

whenever two of the v 's are equal. We denote by $\mathcal{E}^p(V)$ the set of all skew-symmetric p -forms on V .

If f, g are p -forms and c a number, we define new p -forms $f + g, cf$ by

$$(f + g)(v_1, \dots, v_p) = f(v_1, \dots, v_p) + g(v_1, \dots, v_p)$$

$$(cf)(v_1, \dots, v_p) = c f(v_1, \dots, v_p).$$

Exercise. Verify that the set $\mathcal{T}^p(V)$, equipped with the two operations defined above, is a vector space. Also verify that $\mathcal{S}^p(V), \mathcal{E}^p(V)$ are linear subspaces of $\mathcal{T}^p(V)$.

Question. What are their dimensions?

By definition,

$$\mathcal{S}^0(V) = \mathcal{E}^0(V) = \mathcal{T}^0(V) = \mathbf{R}.$$

So they are all one dimensional.

A 1-form is a linear map $f : V \rightarrow \mathbf{R}$, $v \mapsto f(v)$. So, by definition,

$$\mathcal{S}^0(V) = \mathcal{E}^0(V) = \mathcal{T}^0(V) = V^*.$$

So they are all $n = \dim V$ dimensional.

Example. Consider the map $f : \mathbf{R}^2 \times \mathbf{R}^2 \rightarrow \mathbf{R}$ defined by

$$f(X, Y) = X \cdot Y.$$

This is a 2-form because the dot product has properties D1-D3 (chapter 2). It is also symmetric, by property D1. But f is not skew-symmetric because $f(X, X) > 0$ when $X \neq O$. More generally, an inner product $\langle \cdot, \cdot \rangle$ on a vector space V defines a symmetric 2-form.

Example. Consider the map $f : \mathbf{R}^2 \times \mathbf{R}^2 \rightarrow \mathbf{R}$ defined by

$$f(X, Y) = \det[X, Y].$$

This is a 2-form and it is skew-symmetric by the four properties of determinant of a matrix (chapter 5).

Example. Let f be a general 2-form. Expanding $f(X, Y)$ using the standard basis $\{E_1, E_2\}$ of \mathbf{R}^2 , we get

$$\begin{aligned} f(X, Y) &= f(x_1E_1 + x_2E_2, y_1E_1 + y_2E_2) \\ &= x_1y_1f(E_1, E_1) + x_1y_2f(E_1, E_2) + x_2y_1f(E_2, E_1) + x_2y_2f(E_2, E_2). \end{aligned}$$

If f is symmetric, then this becomes

$$f(X, Y) = x_1y_1f(E_1, E_1) + (x_1y_2 + x_2y_1)f(E_1, E_2) + x_2y_2f(E_2, E_2).$$

Example. Continuing with the preceding discussion, we consider the case when f is skew-symmetric. Then

$$(*) \quad f(X, Y) = x_1y_2f(E_1, E_2) + x_2y_1f(E_2, E_1).$$

For $Y = X$, this becomes

$$f(X, X) = x_1x_2(f(E_1, E_2) + f(E_2, E_1)).$$

But since $f(X, X) = 0$ for arbitrary X , it follows that $f(E_1, E_2) = -f(E_2, E_1)$. Hence (*) now reads

$$f(X, Y) = (x_1y_2 - x_2y_1)f(E_1, E_2) = f(E_1, E_2)\det[X, Y]$$

for any X, Y . Thus, f is completely determined by the value of $f(E_1, E_2)$.

Exercise. Expand a skew-symmetric 2-form f on \mathbf{R}^3 and write $f(X, Y)$ in terms of $f(E_1, E_2)$, $f(E_1, E_3)$, $f(E_2, E_3)$.

This chapter focuses primarily on the study of skew-symmetric p -forms. Symmetric 2-forms arise also in eigenvalue problems, as will be seen in the next chapter.

We now make four important observations about general p -forms. Let $\{u_1, \dots, u_n\}$ be a basis of V . First, let f be a p -form. Consider $f(v_1, \dots, v_p)$ for given elements v_1, \dots, v_p in V . Each v_i is a linear combination, say

$$v_i = \sum_{j=1}^n a_{ij}u_j.$$

By linearity, we can expand $f(v_1, \dots, v_p)$ and express it as a sum of the form

$$f(v_1, \dots, v_p) = \sum a_{1j_1} \cdots a_{pj_p} f(u_{j_1}, \dots, u_{j_p})$$

where the sum is ranging over all j_1, \dots, j_p taking values in the list $1, \dots, n$. This shows that f is determined by the values $f(u_{j_1}, \dots, u_{j_p})$. In other words,

1. if f, g are two p -forms such that

$$f(u_{j_1}, \dots, u_{j_p}) = g(u_{j_1}, \dots, u_{j_p})$$

for all j_1, \dots, j_p , then $f = g$.

In particular, if f is a p -form with $f(u_{j_1}, \dots, u_{j_p}) = 0$ for all j_1, \dots, j_p , then f is the zero p -form. (ie. f maps every p -tuple (v_1, \dots, v_p) to 0).

2. p -form f is skew-symmetric iff

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

whenever the p -tuple $x' = (v'_1, \dots, v'_p)$ is obtained from $x = (v_1, \dots, v_p)$ by swapping two of the v 's.

Exercise. Prove this. (Hint: Imitate the proof of skew-symmetry for the determinant in chapter 5.)

Third, let f be a skew-symmetric p -form. Then $f(u_{j_1}, \dots, u_{j_p}) = 0$ whenever two of the j 's are equal. Thus to determine f it is enough to know the values $f(u_{j_1}, \dots, u_{j_p})$ for all j_1, \dots, j_p distinct. In particular when $p > n$, since there cannot be p distinct integers between 1 and n , it follows that $f(u_{j_1}, \dots, u_{j_p}) = 0$ for all j_1, \dots, j_p . Thus when $p > n$, the only skew-symmetric p -form is the zero p -form. Now suppose that $p \leq n$ and that j_1, \dots, j_p are distinct. By the second observation above, if we perform a series of m swaps – swapping two elements at a time – on $x = (u_{j_1}, \dots, u_{j_p})$ to result in, say $x' = (u_{j'_1}, \dots, u_{j'_p})$, then

$$f(x') = (-1)^m f(x).$$

By the Sign Theorem (chapter 5), we can write this as

$$f(x') = \text{sign}(L)f(x)$$

where L is any permutation of p letters which transforms (j_1, \dots, j_p) to (j'_1, \dots, j'_p) . In particular, given any $x' = (u_{j'_1}, \dots, u_{j'_p})$ such that j'_1, \dots, j'_p are distinct, we can transform this by a permutation L so that the result $x = (u_{j_1}, \dots, u_{j_p})$ is such that $1 \leq j_1 < \dots < j_p \leq n$, and that

$$f(x') = \text{sign}(L)f(x).$$

Thus we conclude that

3. a skew-symmetric p -form f is determined by the values

$$f(u_{j_1}, \dots, u_{j_p}), \quad 1 \leq j_1 < \dots < j_p \leq n.$$

Fourth, given a p -tuple $J = (j_1, \dots, j_p)$ of integers such that $1 \leq j_1 < \dots < j_p \leq n$, we can define a p -form f_J as follows. We introduce some notations first. Given any p -tuple $J' = (j'_1, \dots, j'_p)$ of integers (not necessarily distinct) in the list $1, \dots, n$, we write

$$u_{J'} = (u_{j'_1}, \dots, u_{j'_p}).$$

We also introduce the symbol

$$\Delta(J, J') = \text{sign}(L)$$

if J' transforms to J by the permutation L . Again, by the Sign Theorem (chapter 5), this is independent of the choice of permutation L which transforms J' to J . If J' does not transform to J , then we set $\Delta(J, J') = 0$. Let v_1, \dots, v_p be elements in V . Then, in a unique way, each v_i is a linear combination, say

$$v_i = \sum_{j=1}^n a_{ij} u_j.$$

We *define* a map $f_J : V \times \dots \times V \rightarrow \mathbf{R}$ by

$$f_J(v_1, \dots, v_p) = \sum a_{1j'_1} \dots a_{pj'_p} \Delta(J, J')$$

where the sum is ranging over all p -tuples $J' = (j'_1, \dots, j'_p)$ of integers in the list $1, \dots, n$. (Compare this with our second observation above).

4. Then f_J has the following properties:

- f_J is a p -form
- f_J is skew-symmetric
- $f_J(u_{J'}) = \Delta(J, J')$ for all J' .

All three properties can be verified in a straightforward way by imitating the proof of the alternating, linearity (additive and scaling), and unity properties of the determinant function \det on $n \times n$ matrices (chapter 5).

Lemma 9.2. For $0 \leq p \leq n$, there are exactly $\binom{n}{p}$ k -tuples (j_1, \dots, j_p) of integers such that $1 \leq j_1 < \dots < j_p \leq n$.

Proof: We prove this by induction on the integer $n+p$. For $n+p=0$, ie. $n=p=0$, there is just one 0-tuple, namely the empty tuple, by definition. Thus our claim holds trivially in this case. Assuming that the claim is true for $n+p < N$, let's consider case $n+p=N$. We separate p -tuples into two kinds as follows.

- A. All p -tuples of the form (j_1, \dots, j_{p-1}, n) , ie. $j_p = n$. Since $j_{p-1} < j_p = n$, it follows that $1 \leq j_1 < \dots < j_{p-1} \leq n-1$. Thus p -tuples of this kind correspond one-to-one with the $(p-1)$ -tuples (j_1, \dots, j_{p-1}) with $1 \leq j_1 < \dots < j_{p-1} \leq n-1$. Since $n-1+p-1 = N-2 < N$, our inductive assumption applies, and the number of such $(p-1)$ -tuples is $\binom{n-1}{p-1}$. It follows that the number of p -tuples of the form (j_1, \dots, j_{p-1}, n) is also $\binom{n-1}{p-1}$.
- B. All p -tuples (j_1, \dots, j_p) such that $j_p < n$. In other words, these are the p -tuples (j_1, \dots, j_p) such that $1 \leq j_1 < \dots < j_p \leq n-1$. Again, since $n-1+p = N-1 < N$, our inductive assumption applies, and the number of such p -tuples is $\binom{n-1}{p}$.

So we conclude that the total number of p -tuples is

$$\binom{n-1}{p-1} + \binom{n-1}{p} = \binom{n}{p}.$$

(See exercise below.) This completes the proof. \square

Exercise. Prove the last identity of binomial coefficients. Use the polynomial identity:

$$(1+t)^{n-1}(1+t) = (1+t)^n.$$

(What are the coefficients of t^p on both sides?)

Theorem 9.3. For $0 \leq p \leq n$, $\dim \mathcal{E}^p(V) = \binom{n}{p}$.

Proof: Let $\{u_1, \dots, u_n\}$ be a basis of V . We will construct a basis of $\mathcal{E}^p(V)$. By the preceding lemma, the number of p -tuples $J = (j_1, \dots, j_p)$ of integers with $1 \leq j_1 < \dots < j_p \leq n$ is $\binom{n}{p}$. By our fourth observation above, for each such J , we have a skew-symmetric p -form f_J such that

$$f_J(u_{J'}) = \Delta(J, J').$$

To prove our assertion, it suffices to show that these p -forms f_J together form a basis of $\mathcal{E}^p(V)$.

- Linear independence: Consider a linear relation

$$\sum_J b_J f_J = O$$

where the sum is ranging over all p -tuples $J = (j_1, \dots, j_p)$ of integers with $1 \leq j_1 < \dots < j_p \leq n$. Fix a $J' = (j'_1, \dots, j'_p)$ with $1 \leq j'_1 < \dots < j'_p \leq n$, but otherwise arbitrary. Evaluating both sides of the linear relation above on the element $u_{J'} = (u_{j'_1}, \dots, u_{j'_p})$, we get

$$\sum_J b_J \Delta(J, J') = 0.$$

Suppose $\Delta(J, J') \neq 0$ for some J in the sum above. Then by definition,

$$\Delta(J, J') = \text{sign}(L)$$

where J' transforms to J by some permutation L . But since we have

$$1 \leq j_1 < \dots < j_p \leq n, \quad 1 \leq j'_1 < \dots < j'_p \leq n,$$

J' can transform to J only if $J' = J$. Thus we can choose L to be the identity permutation, and hence

$$\Delta(J, J') = 1.$$

Moreover, the linear relation now reads

$$b_{J'} = 0.$$

Since J' is arbitrary, it follows that $b_J = 0$ for all J . This shows that the f_J forms a linearly independent set in $\mathcal{E}^p(V)$.

• **Spanning property:** Let f be any skew-symmetric p -form. We will show that f is a linear combination of the f_J 's. Define a skew-symmetric p -form by

$$g = \sum_J f(u_J) f_J$$

where the sum is ranging over all p -tuples $J = (j_1, \dots, j_p)$ of integers with $1 \leq j_1 < \dots < j_p \leq n$. Again fix a $J' = (j'_1, \dots, j'_p)$ with $1 \leq j'_1 < \dots < j'_p \leq n$, but otherwise arbitrary. Then

$$g(u_{J'}) = \sum_J f(u_J) f_J(u_{J'}) = \sum_J f(u_J) \Delta(J, J').$$

Again, by the very same argument as above, we conclude that $\Delta(J, J') = 0$ unless $J = J'$, and that $\Delta(J', J') = 1$. Thus we get

$$g(u_{J'}) = f(u_{J'}).$$

Since J' is arbitrary, by our third observation above, we conclude that $g = f$. Hence

$$f = \sum_J f(u_J) f_J.$$

This completes our proof. \square

Given a linear map $L : V \rightarrow V$, we define a map

$$L_p : \mathcal{E}^p(V) \rightarrow \mathcal{E}^p(V)$$

for each $p = 0, 1, \dots, n$, as follows. Put $L_0 = 1$ (note that $\mathcal{E}^0(V) = \mathbf{R}$, by definition). For $p \geq 1$, given a p -form f , we define $L_p(f)$ by

$$[L_p(f)](v_1, \dots, v_p) = f[L(v_1), \dots, L(v_p)].$$

We will show that L_p is linear. Let f_1, f_2, f be elements of $\mathcal{E}^p(V)$, and c be a scalar. Let v_1, \dots, v_p be elements of V . Then

$$\begin{aligned} [L_p(f_1 + f_2)](v_1, \dots, v_p) &= (f_1 + f_2)[L(v_1), \dots, L(v_p)] \\ &= f_1[L(v_1), \dots, L(v_p)] + f_2[L(v_1), \dots, L(v_p)] \\ &= [L_p(f_1)](v_1, \dots, v_p) + [L_p(f_2)](v_1, \dots, v_p) \\ &= [L_p(f_1) + L_p(f_2)](v_1, \dots, v_p). \end{aligned}$$

It follows that

$$L_p(f_1 + f_2) = L_p(f_1) + L_p(f_2).$$

Similarly, we have

$$L_p(cf) = cL_p(f).$$

Corollary 9.4. *For a given linear map $L : V \rightarrow V$, the linear map $L_n : \mathcal{E}^n(V) \rightarrow \mathcal{E}^n(V)$ is a scalar multiple of the identity map.*

Proof: Recall that (an exercise in chapter 8) if U is a 1 dimensional vector space U , then any linear map from U to U is a scalar multiple of the identity map on U . By the preceding theorem, the vector space $\mathcal{E}^n(V)$ has dimension $\binom{n}{n} = 1$. It follows that the linear map L_n is a scalar multiple of the identity map on $\mathcal{E}^n(V)$. \square

9.2. The determinant

Definition 9.5. For a given linear map $L : V \rightarrow V$, we define the determinant, denoted by $\det(L)$, to be the number such that $L_n = \det(L)Id$.

Theorem 9.6. (*Multiplicative property*) If $L : V \rightarrow V$ and $M : V \rightarrow V$ are two linear maps, then

$$\det(L \circ M) = \det(L)\det(M).$$

Proof: Let f be an element in $\mathcal{E}^n(V)$, and v_1, \dots, v_p be elements in V . Then

$$\begin{aligned} [(L \circ M)_n f](v_1, \dots, v_n) &= f[(L \circ M)(v_1), \dots, (L \circ M)(v_n)] \\ &= f[L(M(v_1)), \dots, L(M(v_n))] \\ &= [L_n(f)][M(v_1), \dots, M(v_n)] \\ &= [M_n(L_n(f))](v_1, \dots, v_n) \\ &= [(M_n \circ L_n)(f)](v_1, \dots, v_n). \end{aligned}$$

This shows that

$$(L \circ M)_n = M_n \circ L_n.$$

Writing this in terms of \det , we get

$$\det(L \circ M)Id = \det(M)\det(L)Id.$$

This proves our assertion. \square

Corollary 9.7. If $L : V \rightarrow V$ is an invertible linear map, then $\det(L) \neq 0$. In this case, $\det(L^{-1}) = 1/\det(L)$.

Proof: $\det(L \circ L^{-1}) = \det(L)\det(L^{-1}) = 1$. \square

Corollary 9.8. If $L : V \rightarrow V$ is a noninvertible linear map, then $\det(L) = 0$.

Proof: Since L is not invertible, it fails to be injective (chapter 8). Let u_1 be a nonzero element such that $L(u_1) = O$. By suitably adjoining elements from V , we can form a basis

$\{u_1, \dots, u_n\}$ of V . Then relative to this basis, we have a skew-symmetric n -form f_I such that

$$f_I(u_1, \dots, u_n) = 1.$$

Since $L(u_1) = O$, we get

$$[L_n(f_I)](u_1, \dots, u_n) = f_I[L(u_1), \dots, L(u_n)] = 0.$$

On the other hand, $L(f_I) = \det(L)f_I$. It follows that $\det(L) = 0$. \square

Example. Let $V = \mathbf{R}^n$. Then a linear map $L : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is represented by an $n \times n$ matrix $A_L = (a_{ij})$, ie.

$$L(E_i) = \sum_j a_{ij} E_j$$

where $\{E_1, \dots, E_n\}$ is the standard basis of \mathbf{R}^n . Consider the special n -tuple of integers $I = (1, \dots, n)$. Relative to the standard basis, we have a skew-symmetric n -form f_I such that

$$f_I(E_1, \dots, E_n) = 1.$$

By definition, we have

$$L_n(f_I) = \det(L)f_I.$$

Evaluating both sides on (E_1, \dots, E_n) , we get

$$[L_n(f_I)](E_1, \dots, E_n) = \det(L).$$

Unwinding the left hand side, we get

$$\begin{aligned} \det(L) &= f_I[L(E_1), \dots, L(E_n)] \\ &= \sum_J a_{1j_1} \cdots a_{nj_n} f_I(E_{j_1}, \dots, E_{j_n}) \end{aligned}$$

where the sum is ranging over all n -tuple $J = (j_1, \dots, j_n)$ of integers in the list $1, \dots, n$. Since f_I is skew-symmetric, the summand is zero unless j_1, \dots, j_n are distinct. When they are distinct, $J = (j_1, \dots, j_n)$ is a rearrangement of $I = (1, \dots, n)$, ie. J is a permutation of n letters, and that

$$f_I(E_{j_1}, \dots, E_{j_n}) = \Delta(I, J) = \text{sign}(J).$$

It follows that

$$\det(L) = \sum_J \text{sign}(J) a_{1j_1} \cdots a_{nj_n}$$

where the sum now is ranging over all permutations J of n letters. This number, by our previous definition in chapter 5, is the determinant of the $n \times n$ matrix $A_L = (a_{ij})$. Thus we have proven

Theorem 9.9. *For a given linear map $L : \mathbf{R}^n \rightarrow \mathbf{R}^n$,*

$$\det(L) = \det(A_L).$$

Question. *How do we compute $\det(L)$ for a given linear map $L : V \rightarrow V$?*

Let $\{u_1, \dots, u_n\}$ be a basis of V , and let $A_L = (a_{ij})$ be the matrix of L relative to this basis. Now relative to the same basis, we have a skew-symmetric n -form f_I such that

$$f_I(u_1, \dots, u_n) = 1.$$

By proceeding the same way as in the case of \mathbf{R}^n above (with u_1, \dots, u_n playing the roles of E_1, \dots, E_n), we find that

$$\det(L) = \det(A_L).$$

Thus we have the following generalization of the preceding theorem:

Theorem 9.10. *Given a linear map $L : V \rightarrow V$ and a basis $\{u_1, \dots, u_n\}$ of V , let A_L be the matrix of L relative to this basis. Then*

$$\det(L) = \det(A_L).$$

9.3. Appendix

The approach to determinants presented above is based on the notion of p -forms. The main theorem which made our definition for $\det(L)$ possible was the fact that $\mathcal{E}^n(V)$ is one dimensional, which took quite a bit of work to establish. We should emphasize that p -forms are of fundamental importance in many areas of mathematics – not just in algebra. Thus introducing it is worth every bit of work. On the other hand, Theorem 9.10 suggests

a more pragmatic way to define determinants. We now discuss this approach, which is less theoretical (perhaps even easier) and sidesteps the use of p -forms. However, this approach seems less elegant because the definition of determinant in this approach involves a choice of basis, and one has to prove that it is independent of the choice at the end.

Let $L : V \rightarrow V$ be a linear map. Let $\{u_1, \dots, u_n\}$ be a basis of V and let $A = (a_{ij})$ (we drop the subscript for convenience) be the matrix of L relative to this basis. We *define*

$$\det(L) = \det(A).$$

We must prove that this definition is independent of the choice basis $\{u_1, \dots, u_n\}$, ie. that if A' is the matrix of L relative to another basis $\{u'_1, \dots, u'_n\}$ of V , then

$$\det(A') = \det(A).$$

Let $R : V \rightarrow \mathbf{R}^n$ be the linear map defined by $R(u_i) = E_i$ (Theorem 8.2). Then R is a linear isomorphism (why?). We also have a new linear map

$$L_R = R \circ L \circ R^{-1} : \mathbf{R}^n \rightarrow \mathbf{R}^n.$$

From

$$L_R(E_i) = R[L(u_i)] = R\left[\sum_j a_{ij}u_j\right] = \sum_j a_{ij}E_j,$$

it follows that for any column vector X in \mathbf{R}^n ,

$$L_R(X) = A^t X$$

Consider the linear isomorphism $S : V \rightarrow \mathbf{R}^n$ defined by $S(u'_i) = E_i$, and the linear map

$$L_S = S \circ L \circ S^{-1} : \mathbf{R}^n \rightarrow \mathbf{R}^n.$$

Let

$$T = S \circ R^{-1} : \mathbf{R}^n \rightarrow \mathbf{R}^n.$$

Then for any column vector X in \mathbf{R}^n ,

$$T(X) = BX$$

where B is some invertible matrix B . We have

$$T \circ L_R \circ T^{-1} = S \circ R^{-1} \circ R \circ L \circ R^{-1} \circ R \circ S^{-1} = S \circ L \circ S^{-1} = L_S.$$

Applying this to any vector X in \mathbf{R}^n , we get

$$BA^tB^{-1}X = (A')^tX.$$

Since X is arbitrary, it follows that

$$BA^tB^{-1} = (A')^t.$$

Thus

$$\det(A')^t = \det(BA^tB^{-1}).$$

By the multiplicative property of *matrix determinant*, we have

$$\det(A')^t = \det(B)\det(A^t)\det(B)^{-1} = \det(A^t).$$

Since the determinant of a matrix is the same as that of its transpose, This completes the proof.

9.4. Homework

1. Use Theorem 9.10 to give another proof of the multiplicative property of determinant. Recall that if $L : V \rightarrow V$, $M : V \rightarrow V$ are two linear maps of a finite dimensional vector space V , then relative to a given basis, we have (Theorem 8.27)

$$A_{M \circ L} = A_L A_M.$$

2. Let V_1, V_2 be finite dimensional vector spaces and $L_1 : V_1 \rightarrow V_1$, $L_2 : V_2 \rightarrow V_2$ be linear maps. Let $V = V_1 \oplus V_2$. Define a third map $L : V \rightarrow V$ by

$$L(v_1, v_2) = (L_1v_1, L_2v_2).$$

- (a) Show that L is linear.
- (b) Show that $\det(L) = \det(L_1)\det(L_2)$.

3. Let A be an $n \times n$ matrix. Consider the linear map

$$L_A : M(n, m) \rightarrow M(n, m), \quad L_A(B) = AB.$$

Prove that $\det(L_A) = \det(A)^m$.

4. Let $L : V \rightarrow V$ be a linear map of a finite dimensional vector space V , and A be the matrix of L relative to a given basis of V . Show that the matrix of the dual map $L^* : V^* \rightarrow V^*$ relative to the dual basis is A^t . Conclude that

$$\det(L^*) = \det(L).$$

5. * *Volume.* Let V be a finite dimensional vector space equipped with an inner product $\langle \cdot, \cdot \rangle$. Let $\{u_1, \dots, u_n\}$ and $\{u'_1, \dots, u'_n\}$ be two orthonormal bases. Let $L : V \rightarrow V$ be the linear map defined by

$$L(u_i) = u'_i, \quad i = 1, \dots, n.$$

- (a) Show that for any vectors v, w in V ,

$$\sum_{j=1}^n \langle v, u'_j \rangle \langle u'_j, w \rangle = \langle v, w \rangle.$$

- (b) Show that the matrix of L relative to $\{u_1, \dots, u_n\}$ is orthogonal.

- (c) Let v_1, \dots, v_n be vectors in V . Let $M, M' : V \rightarrow V$ be the linear maps defined by

$$M(u_i) = v_i = M'(u'_i), \quad i = 1, \dots, n.$$

Prove that

$$\det(M) = \pm \det(M').$$

- (d) Define the volume of the parallelepiped P generated by v_1, \dots, v_n in V to be

$$\text{Vol}(P) = |\det(M)|.$$

Conclude that this is well-defined, ie. it is independent of the choice of orthonormal basis of V .

6. ** *Another way to define determinant, assuming the notion of a ring.* Let V be a finite dimensional vector space of dimension n . Put

$$TV = \bigoplus_{p \geq 0} T^p V$$

where T^pV is the p -fold tensor product space $V \otimes \cdots \otimes V$ (p times.) Note that $T^0V = \mathbf{R}$ by definition.

(a) Show that TV is naturally a ring with a unique bilinear product $TV \times TV \rightarrow TV$ such that

$$((v_1 \otimes \cdots \otimes v_p), (u_1 \otimes \cdots \otimes u_q)) \mapsto v_1 \otimes \cdots \otimes v_p \otimes u_1 \otimes \cdots \otimes u_q.$$

(b) Let $f : V \rightarrow V$ be a linear map. Show that there is a unique linear map $Tf : TV \rightarrow TV$ such that

$$Tf(v_1 \otimes \cdots \otimes v_p) = fv_1 \otimes \cdots \otimes fv_p.$$

(c) Let AV be the two-sided ideal of the ring TV generated by the set $\{v \otimes v | v \in V\}$. Note that AV is a linear subspace of TV (why?) Show that Tf preserves AV , i.e.

$$Tf(AV) \subset AV.$$

(d) Put

$$\wedge V = TV/AV$$

and let $\wedge^p V$ be the subspace of $\wedge V$ spanned by all cosets of the form

$$v_1 \wedge \cdots \wedge v_p := v_1 \otimes \cdots \otimes v_p + AV.$$

Show that $\wedge^n V$ is one dimensional. In fact, if v_1, \dots, v_n form a basis of V , then $v_1 \wedge \cdots \wedge v_n$ forms a basis of $\wedge^n V$.

(e) Use (b) and (c) to show that there is a unique linear map

$$\wedge f : \wedge V \rightarrow \wedge V$$

such that

$$\wedge f(\wedge v_1 \wedge \cdots \wedge v_p) = fv_1 \wedge \cdots \wedge fv_p.$$

(f) Use (d) to show that $\wedge f$ on $\wedge^n V$ is a scalar, and show that the scalar is precisely

$$\det(f).$$

10. Eigenvalue Problems For Linear Maps

Throughout this chapter, unless stated otherwise, $L : V \rightarrow V$ will be a linear map of a finite dimensional vector spaces V . We denote $\dim V$ by n .

10.1. Eigenvalues

Definition 10.1. *A number λ is called an eigenvalue of L if the matrix $L - \lambda Id_V$ is singular, ie. not invertible.*

The statement that λ is an eigenvalue of L can be restated in any one of the following equivalent ways:

- (a) the linear map $L - \lambda Id_V$ is singular.
- (b) the function $\det(L - x Id_V)$ vanishes at $x = \lambda$.
- (c) $\text{Ker}(L - \lambda Id_V)$ is not the zero space. This subspace is called the eigenspace for λ .
- (d) There is a nonzero vector v such that $L(v) = \lambda v$. Such a *nonzero* vector is called an eigenvector for λ .

Let $A_L = (a_{ij})$ be the matrix of L relative to a given basis. Then for any number λ , the matrix of the linear map $L - \lambda Id_V$ relative to this basis is $A_L - \lambda I$, where I is the $n \times n$ identity matrix. By Theorem 9.10,

$$\det(L - \lambda Id_V) = \det(A_L - \lambda I).$$

This shows the linear map L has the same eigenvalues as its matrix A_L .

Definition 10.2. *The polynomial function $\det(L - xId_V) = \det(A_L - xI)$ is called the characteristic polynomial of L .*

Thus the characteristic polynomial is a polynomial function whose leading term is $(-1)^n x^n$ (chapter 6).

Exercise. Compute the characteristic polynomial of the linear map

$$L : M(2, 2) \rightarrow M(2, 2), \quad A \mapsto BA$$

where $B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$.

Exercise. Do the same for the linear map

$$D : U \rightarrow U, \quad D(f) = f'$$

where $U = \text{Span}\{1, t\}$.

Exercise. Do the same for the linear map

$$D : U \rightarrow U, \quad D(f) = f'$$

where $U = \text{Span}\{\cos t, \sin t\}$.

Theorem 10.3. *Let u_1, \dots, u_k be eigenvectors of L . If their corresponding eigenvalues are distinct, then the eigenvectors are linearly independent.*

Corollary 10.4. *There are no more than n distinct eigenvalues for L .*

The proofs of these two statements are similar to the case of matrices (chapter 6).

10.2. Diagonalizable linear maps

Definition 10.5. A linear map $L : U \rightarrow U$ is said to be diagonalizable if U has a basis consisting of eigenvectors of L . Such a basis is called an eigenbasis of L .

Theorem 10.6. L is diagonalizable iff its matrix A_L relative to some given basis is diagonalizable.

Proof: Let $A = A_L = (a_{ij})$ be the matrix of L relative to a basis $\{u_1, \dots, u_n\}$ of U . Let $f : U \rightarrow \mathbf{R}^n$ be the linear isomorphism defined by

$$f(u_i) = E_i.$$

Then

$$\begin{aligned} (A^t \circ f)(u_i) &= \text{row } i \text{ of } A = \sum_j a_{ij} E_j = \sum_j a_{ij} f(u_j) \\ &= f\left[\sum_j a_{ij} u_j\right] = f[L(u_i)] = (f \circ L)(u_i). \end{aligned}$$

So the two linear maps $A^t \circ f, f \circ L : U \rightarrow \mathbf{R}^n$ agree on a basis, so they are equal:

$$(*) \quad A^t \circ f = f \circ L.$$

Likewise, we have

$$(**) \quad f^{-1} \circ A^t = L \circ f^{-1}.$$

If L is diagonalizable with eigenbasis $\{v_1, \dots, v_n\}$ and corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, then applying $(*)$ to v_i , we find

$$A^t f(v_i) = f[L(v_i)] = \lambda_i f(v_i).$$

This shows that $f(v_1), \dots, f(v_n)$ form an eigenbasis of A^t .

Conversely, if A^t is diagonalizable with eigenbasis $\{B_1, \dots, B_n\}$ and corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, then applying $(**)$ shows that $f^{-1}(B_1), \dots, f^{-1}(B_n)$ form an eigenbasis of L .

Thus we have proved that L is diagonalizable iff A^t is. Now in chapter 6, we showed that A^t is diagonalizable iff A is. \square

10.3. Symmetric linear maps

In this section, $\langle \cdot, \cdot \rangle$ will denote an inner product on U .

Definition 10.7. A linear map $L : U \rightarrow U$ is called a symmetric linear map if $\langle Lv, w \rangle = \langle v, Lw \rangle$ for all $v, w \in U$.

Example. Recall that given an $n \times n$ matrix A , we have the linear map $L_A : \mathbf{R}^n \rightarrow \mathbf{R}^n$, $L_A(X) = AX$. Let $\langle \cdot, \cdot \rangle$ be the dot product. For any vectors X, Y , we have

$$L_A(X) \cdot Y = AX \cdot Y = Y \cdot AX.$$

Since

$$X \cdot L_A(Y) = X^t AY = Y^t A^t X = Y \cdot A^t X$$

it follows that the linear map L_A is symmetric iff $A = A^t$, ie. iff A is a symmetric matrix.

Example. Let A be a symmetric positive definite $n \times n$ matrix (see homework section of chapter 7). Then we have an inner product on \mathbf{R}^n defined by

$$\langle X, Y \rangle = X \cdot AY.$$

Let B be a symmetric $n \times n$ matrix, and consider the linear map L_B defined by $L_B(X) = BX$. Note that the matrix of L_B relative to the standard basis of \mathbf{R}^n is B^t . Now

$$\langle L_B X, Y \rangle = BX \cdot AY = X^t B^t AY = X \cdot BAY.$$

We also have

$$\langle X, L_B Y \rangle = X \cdot ABY.$$

It follows that the linear map L_B is symmetric (with respect to the inner product $\langle \cdot, \cdot \rangle$) iff $BA = AB$.

Exercise. Given an example to show that a linear map L need not be symmetric even if its matrix relative to some given basis is symmetric.

Given a linear map L , we define a map $g_L : U \times U \rightarrow \mathbf{R}$ by

$$g_L(v_1, v_2) = \langle L(v_1), v_2 \rangle.$$

Exercise. Verify that g_L is a 2-form on U . Moreover, g_L is a symmetric 2-form iff L is a symmetric linear map. A symmetric 2-form is also called a *quadratic form*. g_L is called the quadratic form associated with L .

Theorem 10.8. *Let g be a 2-form on U . Then there is a unique linear map $L : U \rightarrow U$ such that $g = g_L$. Moreover, g is symmetric iff L is symmetric.*

Proof: The second assertion follows immediately from the first assertion and the preceding exercise. We will construct a linear map L with $g = g_L$, and prove that such an L is unique.

- Construction. Define a map $S : U \rightarrow U^*$, $u \mapsto h_u$, by

$$h_u(v) = g(u, v).$$

Note that since g is linear in the second slot, it follows that $h_u : U \rightarrow \mathbf{R}$ is linear. The map S is linear (exercise). Recall that there is an linear isomorphism

$$R : U \rightarrow U^* = \text{Hom}(U, \mathbf{R}), \quad u \mapsto f_u$$

where f_u is defined by $f_u(v) = \langle u, v \rangle$ (homework section, chapter 8). Consider the linear map

$$L = R^{-1} \circ S : U \rightarrow U.$$

For any element u in U , we have

$$R[L(u)] = S(u), \quad \text{ie. } f_{L(u)} = h_u.$$

Since both sides are elements of U^* , this is equivalent to saying that, for any element v in U ,

$$f_{L(u)}(v) = h_u(v), \quad \text{ie. } g_L(u, v) = \langle L(u), v \rangle = g(u, v).$$

Since this holds for arbitrary u, v , we conclude that $g_L = g$.

- Uniqueness. Suppose L, L' are two linear maps with $g_L = g_{L'}$. Then for any u, v , we have

$$\langle L(u), v \rangle = \langle L'(u), v \rangle, \quad \text{ie. } \langle L(u) - L'(u), v \rangle = 0.$$

Since v is arbitrary, it follows that $L(u) - L'(u) = O$. But u is also arbitrary. Thus $L = L'$. This completes the proof. \square

Theorem 10.9. *L is symmetric iff its matrix A_L relative to any orthonormal basis is symmetric.*

Proof: Let $A_L = (a_{ij})$ be the matrix of L relative to a basis $\{u_1, \dots, u_n\}$ of U (relative to a given inner product $\langle \cdot, \cdot \rangle$). Then

$$\langle u_i, L(u_j) \rangle = \langle u_i, \sum_k a_{jk} u_k \rangle = a_{ji}$$

since $\langle u_i, u_k \rangle = 0$ unless $i = k$, and $\langle u_i, u_i \rangle = 1$. Similarly, we have

$$\langle L(u_i), u_j \rangle = a_{ij}.$$

If L is symmetric, then

$$(*) \quad \langle u_i, L(u_j) \rangle = \langle L(u_i), u_j \rangle$$

for all i, j , and hence A_L is symmetric. Conversely, suppose that A_L is symmetric. Then $(*)$ holds for all i, j . Multiplying both sides by arbitrary numbers x_i and y_j , and then sum over i, j , we get from $(*)$

$$\langle x, L(y) \rangle = \langle L(x), y \rangle$$

where $x = \sum_i x_i u_i$ and $y = \sum_j y_j u_j$. This holds for arbitrary elements x, y of U . It follows that L is symmetric. \square

Theorem 10.10. *Every symmetric map is diagonalizable.*

Proof: Let L be a symmetric map. By the preceding theorem, its matrix A_L relative some orthonormal basis is symmetric, hence is also diagonalizable (chapter 6). By Theorem 10.6, it follows that L is diagonalizable. \square

10.4. Homework

1. Find the characteristic polynomial, eigenvalues, and eigenvectors of the linear map $D^2 + 2D : U \rightarrow U$, where $U = \text{Span}\{e^t, te^t\}$ and $D(f) = f'$.

2. Let $S(2)$ be the space of 2×2 symmetric matrices, and let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Define the linear map

$$L : S(2) \rightarrow S(2), \quad L(B) = A^t B A.$$

Find the characteristic polynomial, eigenvalues, and eigenvectors of L .

3. Repeat the previous problem, but with $S(2)$ replaced by the space of all 2×2 matrices.
4. Let L, M be two linear maps on a finite dimensional vector space U . Show that $L \circ M$ and $M \circ L$ have the same eigenvalues.
5. Let $L : U \rightarrow U$ be a linear map and $\{u_1, \dots, u_n\}$ be an eigenbasis of L with distinct eigenvalues. Show that if u is an eigenvector of L , then u is a scalar multiple of a u_i .
6. Let U be a finite dimensional vector space. Two linear maps L, M on U are said to be *similar* if

$$L = P^{-1} \circ M \circ P$$

for some linear map P on U . Show that similar linear maps have the same eigenvalues.

11. Jordan Canonical Form

Up to now, we've been working with notions involving only real numbers \mathbf{R} . They have been the only “scalars” we use. Vectors in \mathbf{R}^n are, by definition, lists of real numbers. We've scaled vectors by real numbers only. Our matrices have only real numbers as their entries. While working with real numbers alone afford us a certain degree of simplicity, it also places certain restrictions on what we can do. For example, a 2×2 real matrix need not have any real eigenvalue. The matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is one such example. So diagonalizing such a matrix using real numbers alone is out of the question. The source of this restriction is simply that certain polynomial functions, such as $t^2 + 1$, have no real roots. Thus for such matrices or linear maps, we lack the useful numerical tool, provided by eigenvalues and eigenspaces, for studying linear maps. It is a lot harder to compare these maps without those numerical tools.

To circumvent this fundamental difficulty, we therefore need another system of scalars in which we are guaranteed to find a root for every polynomial function, such as $t^2 + 1$. Such systems exist, and the simplest of them is the system of complex numbers.

11.1. Complex numbers

Recall that \mathbf{C} is simply the set \mathbf{R}^2 , endowed with the rules of addition and multiplication given by

$$(x, y) + (x', y') = (x + x', y + y'), \quad (x, y)(x', y') = (xx' - yy', xy' + yx').$$

If we put $z = (x, y)$, then we say that x is the real part of z and y is the imaginary part of z , and we write

$$x = \operatorname{Re}(z), \quad y = \operatorname{Im}(z).$$

We also write $\bar{z} = (x, -y)$ and call this the complex conjugate of z . It is not obvious, but quite easy to verify directly, that the rules of addition and multiplication of \mathbf{C} have all the usual properties that one might expect of scalars. For example, if a, b, c are three complex numbers, then we have the relations

$$ab = ba, \quad (a + b) + c = a + (b + c), \quad (ab)c = a(bc).$$

(Prove this!) It is customary to write $z = x(1, 0) + y(0, 1)$ simply as $z = x + yi$, where $i = (0, 1)$. Note that the complex number $(1, 0)$ does indeed behave like a “one” because $(x, y)(1, 0) = (x, y)$ according to our rule of multiplication. So there is no confusion in writing $(1, 0) = 1$. We can, of course, think of a real number x as a complex number with $\operatorname{Im}(x) = 0$. We also have $i^2 = ii = -(1, 0) = -1$. Note that i is, therefore, precisely a root for the polynomial function $t^2 + 1$, because $i^2 + 1 = 0$. More generally, we have

Theorem 11.1. *(The fundamental theorem of algebra) Every non-constant polynomial function $a_n t^n + \cdots + a_1 t + a_0$, with coefficients $a_i \in \mathbf{C}$, has at least one root in \mathbf{C} .*

This is a statement about certain algebraic completeness of \mathbf{C} which the reals \mathbf{R} lack.

Corollary 11.2. *Every non-constant polynomial function $f(t)$ with coefficients in \mathbf{C} can be factorized into a product of linear functions. Specifically, if $f(t)$ has degree n , i.e. the leading term of $f(t)$ is at^n with $a \neq 0$, then $f(t)$ is the product of n linear factors and a scalar, i.e. $f(t) = a(t - \lambda_n) \cdots (t - \lambda_1)$ where $\lambda_1, \dots, \lambda_n$ are the roots of $f(t)$.*

Proof: By the fundamental theorem of algebra, $f(t)$ has at least one root, say $\lambda_n \in \mathbf{C}$. By long division, we can divide $f(t)$ by $t - \lambda_n$ and the remainder must be a scalar, say b . So we have

$$f(t) = g_1(t)(t - \lambda_n) + b$$

where $g_1(t)$ is a polynomial function with leading term at^{n-1} . Since both sides must be zero when $t = \lambda_n$, it follows that $b = 0$. If $g_1(t)$ is constant, i.e. if $n = 1$, we are done. If not, we

can apply the fundamental theorem to $g_1(t)$ and get a factorization $g_1(t) = g_2(t)(t - \lambda_{n-1})$ where $g_2(t)$ has leading term at^{n-2} . So we get

$$f(t) = g_2(t)(t - \lambda_n)(t - \lambda_{n-1}).$$

This procedure terminates after n steps and we get the desired factorization for $f(t)$. \square

Note that the roots $\lambda_1, \dots, \lambda_n$ of $f(t)$ above need not be distinct. It can be shown that the factorization $f(t) = a(t - \lambda_n) \cdots (t - \lambda_1)$ is *unique* up to rearranging that roots. That is, if $f(t) = b(t - \lambda'_n) \cdots (t - \lambda'_1)$, then $b = a$ and we can rearrange $\lambda'_1, \dots, \lambda'_n$ so that $\lambda_i = \lambda'_i$ for all i .

11.2. Linear algebra over \mathbf{C}

Almost all of the linear algebra notions and results we've developed in the previous chapters when \mathbf{R} is the underlying scalar system, can be developed in a parallel fashion when \mathbf{R} is replaced by \mathbf{C} as the underlying scalars. We mention some of them here and point out a few things that need to be handled with care when working over \mathbf{C} . We will leave some of the details to the reader in the form of exercises at the end of this chapter.

- Row operations on complex matrices, i.e. matrices with entries in \mathbf{C} , can be performed just as over \mathbf{R} .
- The notion of inner product of two vectors $A = (a_1, \dots, a_n), B = (b_1, \dots, b_n)$ in \mathbf{C}^n now becomes $\langle A, B \rangle = a_1 \bar{b}_1 + \cdots + a_n \bar{b}_n$; length becomes $\|A\| = \sqrt{\langle A, A \rangle}$. Note that $\langle A, A \rangle \geq 0$ by definition. It is important to note that while the dot product $X \cdot Y$ in \mathbf{R}^n is linear with respect to scaling of X, Y by *real* numbers, $\langle A, B \rangle$ for complex vectors A, B is only linear with respect to scaling A , and is anti-linear with respect to scaling B . That is, we have

$$\langle zA, B \rangle = z\langle A, B \rangle, \quad \langle A, zB \rangle = \bar{z}\langle A, B \rangle.$$

There is another important difference between the real and the complex case. While $X \cdot Y = Y \cdot X$ for real vectors, we have

$$\langle A, B \rangle = \overline{\langle B, A \rangle}$$

for complex vectors.

- Schwarz's inequality becomes

$$|\langle A, B \rangle| \leq \|A\| \|B\|.$$

- Rules of scaling, adding, and multiplying complex matrices are the same as for real matrices. Likewise for the rules of inverting and transposing matrices, as well as the correspondence between matrices and linear transformations of \mathbf{C}^n .

- The notions of linear subspaces of \mathbf{R}^n , linear relations, linear dependence, bases, and dimension, all carry over to \mathbf{C}^n verbatim. Orthonormal bases and Gram-Schmidt also carry over, but with $\langle A, B \rangle$ playing the role of dot product.

- The notion of the determinant of an $n \times n$ matrix and its basic properties also carry over verbatim.

- But now, unlike real matrices, every $n \times n$ complex matrix has at least one eigenvalue, by the fundamental theorem of algebra.

- Unlike symmetric real matrices, symmetric complex matrices need not be diagonalizable.

E.g. $\begin{bmatrix} 1+i & 1 \\ 1 & 1-i \end{bmatrix}$.

- There is a more subtle complex analogue of the transpose of a matrix, called the conjugate transpose. If a_{ij} is the (ij) entry of A , then the conjugate transpose A^\dagger is the matrix whose (ij) entry is \bar{a}_{ji} . A matrix A is called hermitian if $A^\dagger = A$. This is the complex analogue of a symmetric matrix. It is a theorem that *every hermitian matrix is diagonalizable*. This is the spectral theorem for hermitian matrices. The proof is essentially parallel to the proof of the spectral theorem for real symmetric matrices.

- The notions of abstract vector spaces over \mathbf{C} are similar to the real case. We have built a theory of abstract (real) vector spaces by using the theory for \mathbf{R}^n as a guide. One can now go one step further by replacing the role of \mathbf{R} everywhere by \mathbf{C} . It turns out that virtually

every formal notion in the real case, studied in Chapters 8-10 (see table of content there), has a straightforward analogue in the complex case. We would encourage the reader to write down these formal analogues yourself, or to consult another textbook for details. We would recommend *S. Lang, Linear Algebra*.

11.3. Similarity

From now on, all matrices are assumed to be complex matrices.

Definition 11.3. We say that two $n \times n$ matrices A, B are similar if there is an invertible matrix C such that $B = CAC^{-1}$. In this case, we also say that A, B are conjugates of each other.

Exercise. Show that if P, Q are similar and if Q, R are similar, then P, R are also similar.

Exercise. Show that if $B = CAC^{-1}$, then $B^k = CA^kC^{-1}$ for all positive integer k .

For positive integers k, l , we denote by $O_{k \times l}$ the $k \times l$ zero matrix.

Definition 11.4. An $n \times n$ matrix B is called a Jordan matrix if it has the shape

$$B = \begin{bmatrix} B_1 & O_{k_1 \times k_2} & & \cdots & O_{k_1 \times k_p} \\ O_{k_2 \times k_1} & B_2 & O_{k_2 \times k_3} & \cdots & O_{k_2 \times k_p} \\ \vdots & \vdots & & \vdots & \\ O_{k_p \times k_1} & & \cdots & O_{k_p \times k_{p-1}} & B_p \end{bmatrix}.$$

where the block B_i is a $k_i \times k_i$ matrix whose diagonal entries are all the same, say λ_i , and whose entries just above the diagonal are all 1, and whose entries are 0 everywhere else. Note that $n = k_1 + \cdots + k_p$. The B_i are called the Jordan blocks of B .

If a given Jordan block B_i above is 1×1 , then $B_i = [\lambda_i]$. If B_i is 2×2 , then $B_i = \begin{bmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{bmatrix}$. If B_i is 3×3 , then $B_i = \begin{bmatrix} \lambda_i & 1 & 0 \\ 0 & \lambda_i & 1 \\ 0 & 0 & \lambda_i \end{bmatrix}$, and so on. In particular, any diagonal matrix is a Jordan matrix whose Jordan blocks all have size 1×1 . Note that the λ_i need not be distinct.

Theorem 11.5. (*Jordan canonical form*) Every given $n \times n$ matrix A is similar to a Jordan matrix B . Moreover, B is uniquely determined by A up to rearranging the blocks B_i along its diagonal. B is called a Jordan canonical form of A .

We will sketch a proof of this in the next section and Appendices A,B, of this chapter, assuming some basic facts about the polynomial algebra $\mathbf{C}[x]$. The preceding theorem can be stated more abstractly as follows.

Theorem 11.6. For any given linear map $A : V \rightarrow V$ of a finite dimensional complex vector space V , there is a basis β of V , such that the matrix $B = [A]_\beta$ of A in the basis β is a Jordan matrix. Moreover, B is unique up to rearranging the Jordan blocks. The β above is called a Jordan basis associated to A .

We now sketch the essential steps of the proof. This will also lead to a simple procedure for computing a Jordan matrix of A .

11.4. Invariant subspaces and cycles

Throughout this section, we let $A : V \rightarrow V$ be a linear map of a finite dimensional complex vector space V of dimension n , and denote by $f(t)$ the characteristic polynomial $\det(A - tI)$ of A .

By the fundamental theorem of algebra, $(-1)^n f(t)$ factorizes into a product of functions of the form $(t - \lambda)^{m_\lambda}$, and there is one such factor for every distinct eigenvalue λ of A . The positive integer m_λ is called the *multiplicity* of λ .

Definition 11.7. A linear subspace K of V is said to be A -invariant if $AK \subset K$, i.e. $Av \in K, \forall v \in K$.

Put

$$K_\lambda = \{v \in V | (A - \lambda I)^p v = 0 \text{ for some } p > 0\}.$$

This linear subspace of V is called the *generalized eigenspace* of A corresponding to λ .

Lemma 11.8. K_λ is a A -invariant.

Proof: The maps A , $A - \lambda I$, commute with each other. So if $v \in K_\lambda$, say $(A - \lambda I)^p v = 0$, then

$$(A - \lambda I)^p Av = A(A - \lambda I)^p v = A0 = 0,$$

and so $Av \in K_\lambda$. This shows that $AK \subset K$. \square

Theorem 11.9. (Generalized eigenspace decomposition) We have $\dim K_\lambda = m_\lambda$ for each eigenvalue λ . Moreover, V is equal to the direct sum of all the K_λ . In other words, if $\lambda_1, \dots, \lambda_s$ are the distinct eigenvalues of A , then we have a direct sum decomposition

$$V = K_{\lambda_1} + \cdots + K_{\lambda_s}.$$

A proof is given in Appendix A of this chapter.

Since $AK_\lambda \subset K_\lambda$, we get a map $K_\lambda \rightarrow K_\lambda$, $v \mapsto Av$. This map is denoted by $A|K_\lambda$ and it is called the restriction of A to K_λ . This theorem tells us that we should break up the problem of finding a Jordan matrix of A into the problems of finding a Jordan matrix of each restriction $A|K_\lambda$, and then we can “stack” these smaller Jordan matrices together to give a Jordan matrix of A . In fact, more generally we have

Lemma 11.10. If V_1, V_2 are A -invariant subspaces of A and if $V = V_1 + V_2$ is a direct sum, then a Jordan matrix of A is given by $B = \begin{bmatrix} B_1 & O \\ O & B_2 \end{bmatrix}$ where B_1, B_2 are Jordan matrices of the restrictions $A|V_1$, $A|V_2$ respectively.

Proof: Exercise.

Thus from now on, we will focus on a fixed eigenvalue λ , and put

$$m = m_\lambda, \quad K = K_\lambda \neq (0),$$

which we assume are given to us. Note that the restriction $A|K$ has characteristic polynomial $(-1)^m(t - \lambda)^m$. Since we are focusing on one eigenvalue at a time, we may as well forget V and just consider the linear map

$$A|K_\lambda : K_\lambda \rightarrow K_\lambda.$$

For convenience, we will temporarily drop the λ from the notations and simply write $A : K \rightarrow K$.

The remaining task *to find a Jordan matrix of the linear map $A : K \rightarrow K$, assumed to have characteristic polynomial $(-1)^m(t - \lambda)^m$, and to prove its uniqueness.* The idea is to further decompose K into smaller A -invariant subspaces, until the pieces can't be decomposed any more.

Throughout the rest of this section, λ is assumed fixed and we put

$$T = A - \lambda I.$$

Note that a subspace L of K is A -invariant iff it is T -invariant, and that for $v \in K$ we have

$$Av = \lambda v \Leftrightarrow Tv = 0 \Leftrightarrow v \in \text{Ker}(T).$$

Definition 11.11. *A T -cycle of size $p \geq 1$ is a list of vectors in K of the shape*

$$J = \{v, Tv, \dots, T^{p-1}v\} \tag{1}$$

where $T^{p-1}v \neq 0$ and $T^pv = 0$. A subspace L of K is called T -cyclic if it is spanned by a T -cycle. It is clear that a T -cyclic subspace is also T -invariant.

Theorem 11.12. *Any T -cycle is linearly independent.*

Proof: We give two proofs. The first proof is by induction. A T -cycle of size one $\{v\}$ with $v \neq 0$ is clearly linearly independent. Suppose our assertion holds for any T -cycle of size up to $p - 1$. We now show that for $p > 1$, (1) is linearly independent. Consider a linear relation

$$a_0v + a_1Tv + \dots + a_{p-1}T^{p-1}v = 0.$$

Applying T to this, we get $a_0Tv + \dots + a_{p-2}T^{p-1}v = 0$ since $T^pv = 0$. But $\{Tv, \dots, T^{p-1}v\}$ is clearly a T -cycle of size $p - 1$. By our inductive hypothesis, this is linearly independent, so $a_0 = \dots = a_{p-2} = 0$. This shows that $a_{p-1}T^{p-1}v = 0$, implying that $a_{p-1} = 0$ as well. So, (1) is linearly independent.

The second proof assumes some basic facts about the polynomial algebra $\mathbf{C}[x]$. Let L be the T -cyclic subspace spanned by (1). Consider the map $\varphi : \mathbf{C}[x] \rightarrow L$ defined by $g(x) \mapsto g(T)v$. It is easy to see that $\text{Ker}(\varphi)$ is an ideal in $\mathbf{C}[x]$, hence the ideal must be principal, i.e. it is generated by a unique monic $f(x) \in \mathbf{C}[x]$. Note that $x^p \in \text{Ker}(\varphi)$, hence $x^p | f(x)$. So $\deg f(x) \geq p$. But since φ is onto, it induces an isomorphism $\varphi : \mathbf{C}[x]/(f(x)) \rightarrow L$. So we have

$$p \leq \dim \mathbf{C}[x]/(f(x)) = \deg f(x) = \dim L \leq p$$

implying that all five terms are equal. This shows that $\dim L = p$, so J must be a basis of L , so it is linearly independent. \square

Theorem 11.13. *There is a direct sum decomposition*

$$K = L_1 + \cdots + L_r$$

where each L_i is a T -cyclic subspace of K .

A proof is given in the Appendix B.

It can be shown that a T -cyclic subspace L is indecomposable in the sense that if L is written as a direct sum of two T -invariant subspaces, then one of them has to be zero. This result won't be proved or needed here. Note also that there may be many ways to decompose K into T -cyclic subspaces. But as shown below, while the L_i appearing in the decomposition need not be unique, the list of their dimensions $\dim L_i$ is uniquely determined by A .

By Theorem 11.12, a T -cyclic subspace spanned by a T -cycle J of size p must have $\dim L = p$. In this case, the matrix of the restriction $A|L : L \rightarrow L$ in the basis J is precisely the $p \times p$ matrix

$$[A|L]_J = \begin{bmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \\ 0 & 0 & \cdots & & 0 & \lambda \end{bmatrix} \quad (2)$$

which has the shape of a Jordan block. Since λ is assumed given, this block is completely specified by its size, namely $p = \dim L$. By Lemma 11.10, Theorem 11.13 implies the existence of a Jordan matrix of $A : K \rightarrow K$ with r Jordan blocks $[A|L_i]_{J_i}$, where J_i is a

T -cycle spanning L_i . Combining this with Theorem 11.9, we get the existence assertion of Theorem 11.6. We shall see below that, assuming existence, the uniqueness assertion can be proved by showing that the block sizes, if ordered decreasingly,

$$(*) \quad \dim L_1 \geq \cdots \geq \dim L_r,$$

are uniquely determined by A . In fact, the numbers $(*)$ will be expressed in terms of the numbers $m_i = \dim \text{Ker}(T^i)$, as shown below.

By Theorem 11.13, the list $(*)$ is subject to the conditions

$$m = \dim K = \dim L_1 + \cdots + \dim L_r, \quad \dim L_i > 0. \quad (3)$$

Let's examine a few simple cases.

If $m = 1$, then there is just one possibility for the list $(*)$, namely, 1. A Jordan matrix of A must be 1×1 , so $[\lambda]$ is a Jordan matrix of A .

If $m = 2$, then there are just two distinct possibilities for the list $(*)$, namely, 2, or 1,1. A Jordan matrix of A in these cases are respectively

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}.$$

If $m = 3$, then there are just three possibilities for the list $(*)$: 3, or 2,1, or 1,1,1. A Jordan matrix of A in these cases are respectively

$$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \quad \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}.$$

As m gets larger, the number of possibilities grows and they correspond to the ways in which we can express m as a sum of positive integers as in (3). Thus to find a correct list $(*)$ corresponding to a given A , it is clearly not enough just to know m, λ . We need to extract more out of the preceding theorem. First, we need a more definitive description of K .

Lemma 11.14. *There exists a smallest positive integer q such that*

$$\text{Ker}(T^q) = \text{Ker}(T^{q+1}).$$

Moreover, we have $K = \text{Ker}(T^q)$.

Proof: Clearly we have $\text{Ker}(T) \subset \text{Ker}(T^2) \subset \cdots \subset K$, and by definition, K is the union of all the $\text{Ker}(T^p)$. Since K is finite dimensional, these subspaces cannot grow indefinitely in dimensions, which means that $\dim \text{Ker}(T^p) = \dim K = m$ for all large enough p . In particular, $\text{Ker}(T^q) = \text{Ker}(T^{q+1})$ for some q . Fix the smallest such q , and our first assertion follows.

To prove our second assertion, it suffices to show that $K \subset \text{Ker}(T^q)$. So let $v \in K$, $v \neq 0$; we will show that $T^q v = 0$. By definition of K , we have $T^p v = 0$ for some $p > 0$. If $p \leq q$, then

$$T^q v = T^{q-p} T^p v = 0$$

hence $v \in \text{Ker}(T^q)$. Now assume $p > q$. Then

$$\begin{aligned} 0 &= T^p v = T^{q+1} T^{p-q-1} v \\ &\implies T^{p-q-1} v \in \text{Ker}(T^{q+1}) = \text{Ker}(T^q) \\ &\implies 0 = T^q T^{p-q-1} v = T^{p-1} v. \end{aligned}$$

This shows that for $p > q$, we can still get $T^p v = 0$ after reducing p by 1. We can continue to do so until $p = q$, proving that $T^q v = 0$. \square

We now begin computing the ordered list (*). What is the largest possible value for $\dim L_1$? By Theorem 11.13, L_1 is spanned by a T -cycle J . Now comes a key observation: this J cannot have size larger than q , given by the preceding lemma. For otherwise, J would contain the vector $T^{p-1} v$ for some $p > q$; but this vector in J would be zero, by the preceding lemma, contradicting Theorem 11.12. So, the largest possible value of $\dim L_1$ is q . In particular, the dimensions in the list (*) all lie between 1 and q . For $1 \leq j \leq q$, let's denote by l_j , the number of times the number j occurs in the list (*). So the list (*), in this notation, has the shape

$$\{q, \dots, q\}, \{q-1, \dots, q-1\}, \dots, \{1, \dots, 1\}.$$

We have grouped the list into q smaller sublists, where the $(q-j+1)$ st sublist reads $\{j, \dots, j\}$ having size l_j . Note that l_j can be zero. Now, our remaining task is to find the numbers l_q, \dots, l_1 .

Exercise. Show that if L is a d dimensional T -cyclic subspace of K , then

$$\dim \text{Ker}(T|L)^j = \begin{cases} j & \text{if } 1 \leq j \leq d \\ d & \text{if } j > d. \end{cases}$$

Exercise. Show that $\text{Ker}(T^j) = \text{Ker}(T|L_1)^j + \cdots + \text{Ker}(T|L_r)^j$, and that the right side is a direct sum.

Lemma 11.15. Let $m_j = \dim \text{Ker}(T^j)$, $j = 1, \dots, q$. Then we have

$$\begin{aligned} m_q &= ql_q + (q-1)l_{q-1} + \cdots + l_1 \\ m_{q-1} &= (q-1)l_q + (q-1)l_{q-1} + \cdots + l_1 \\ m_{q-2} &= (q-2)l_q + (q-2)l_{q-1} + (q-2)l_{q-2} + \cdots + l_1 \\ &\dots \\ m_1 &= l_q + l_{q-1} + \cdots + l_1. \end{aligned}$$

Proof: This is an immediate consequence of the preceding two exercises. \square

Example 11.16.

Suppose $A : K \rightarrow K$ has $q = 2$, $m_2 = m = \dim K = 6$, $m_1 = 4$. Then the preceding lemma yields

$$m_2 = 2l_2 + l_1, \quad m_1 = l_2 + l_1.$$

Solving them gives $l_2 = 2$, $l_1 = 2$. So a Jordan matrix of A in this case has two Jordan blocks of size 2 and two Jordan blocks of size 1:

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

where all the unfilled entries are 0.

Exercise. Show that

$$\begin{aligned} l_q &= m_q - m_{q-1} \\ l_q + l_{q-1} &= m_{q-1} - m_{q-2} \\ l_q + l_{q-1} + l_{q-2} &= m_{q-2} - m_{q-3} \\ &\dots \\ l_q + l_{q-1} + \dots + l_1 &= m_1 - 0. \end{aligned}$$

Hence conclude that the m_j determines the l_j .

Thus we have proved

Theorem 11.17. *If $K = L_1 + \dots + L_r$ is a direct sum, where the L_i are nonzero T -cyclic subspaces of K such that*

$$(*) \quad \dim L_1 \geq \dots \geq \dim L_r,$$

then the numbers $()$ are uniquely determined by T (hence by $A = T + \lambda I$.)*

Thus to complete our task of finding a Jordan matrix of A , it is enough to compute the number q , and the numbers m_1, \dots, m_q . Then solve for the numbers l_1, \dots, l_q by the preceding lemma, and we are done. Recall that if we know a matrix of $T : K \rightarrow K$ (in any given basis of K), then computing $m_j = \dim \text{Ker}(T^j)$ can be easily done by row reducing the matrix of T^j . Note also that the number q of Lemma 11.14 can be computed at the same time. Namely, q is the first positive integer such that T^q and T^{q+1} have the same rank ($\dim K = \dim \text{Ker} + \text{rank}$), which can also be easily computed by row reduction.

Recapitulating what we have done, we summarize the steps for finding a Jordan matrix for a given linear map $A : V \rightarrow V$ schematically as follows:

$$V \rightarrow K_\lambda \rightarrow L_i \rightarrow \dim L_i \rightarrow \dim \text{Ker}(A - \lambda I)^j \rightarrow \text{row reduce } T^j = (A - \lambda I)^j.$$

Exercise. Suppose $A : K_\lambda \rightarrow K_\lambda$ has $q = 2$, $m = m_2 = 5$, $m_1 = 2$. Find a Jordan matrix of A .

Exercise. Suppose that a linear map $A : V \rightarrow V$ has characteristic polynomial $-(t - \lambda_1)^2(t - \lambda_2)^3$ where $\lambda_1 \neq \lambda_2$, and that its generalized eigenspaces are

$$K_{\lambda_1} = \text{Ker}(A - \lambda_1 I)^2 \supsetneq \text{Ker}(A - \lambda_1 I), \quad K_{\lambda_2} = \text{Ker}(A - \lambda_2 I)^2 \supsetneq \text{Ker}(A - \lambda_2 I).$$

Find a Jordan matrix of A .

11.5. Appendix A

We now prove Theorem 11.9. Throughout this section, $A : V \rightarrow V$ will be a linear map of a finite dimensional vector space V of dimension n . Our proof involves the use of the notion of the quotient of V by a linear subspace K . This notion is used in Appendix B as well. Let m be the dimension of K .

Define an equivalence relation on the set V as follows. We say that $u, v \in V$ are equivalent if $u - v \in K$. (Check that this is an equivalence relation!) Let V/K denote the set of equivalence classes. The equivalence class containing $v \in V$ is denoted by $\bar{v} = v + K$. If $x \in \mathbf{C}$ and $v \in V$, we define *vector scaling*

$$x\bar{v} = \overline{xv}.$$

This is well-defined; for if u, v represents the same equivalence class, then $xu - xv = x(u - v) \in K$, because K is a linear subspace, and so $\overline{xu} = \overline{xv}$. If $u, v \in V$, we define *vector addition*

$$\bar{u} + \bar{v} = \overline{u + v}.$$

Again, it is easy to check that this is well-defined. With the two operations we have just defined on the set V/K , it is straightforward, but tedious, to check that V/K becomes a complex vector space with the zero vector $\bar{0}$. Moreover, the map

$$V \rightarrow V/K, \quad v \mapsto \bar{v},$$

is linear and surjective. It is called the *projection map*.

Suppose now that K is A -invariant. Define

$$\bar{A} : V/K \rightarrow V/K, \quad \bar{v} = \overline{Av}.$$

Again, this is well-defined. It is also linear, i.e.

$$\bar{A}(\bar{u} + x\bar{v}) = \bar{A}\bar{u} + x\bar{A}\bar{v}.$$

The linear map \bar{A} is called the map induced by A on the quotient, or simply the *induced map*.

Lemma 11.18. *Let $T : V \rightarrow V$ be a linear map, and K be a T -invariant subspace of V . Then we have*

$$\det(T) = \det(T|K) \det(\bar{T}).$$

Proof: Pick a basis of K , say $\{v_1, \dots, v_m\}$. Extend this to a basis of V , say $\beta = \{v_1, \dots, v_m, \dots, v_n\}$. Then Tv_i takes the form

$$Tv_i = \begin{cases} \sum_{j=1}^m p_{ij}v_j & 1 \leq i \leq m \\ \sum_{j=1}^m q_{ij}v_j + \sum_{j=m+1}^n r_{ij}v_j & m+1 \leq i \leq n \end{cases}$$

where $P = (p_{ij}), Q = (q_{ij}), R = (r_{ij})$ are matrices of respective sizes $m \times m, m \times (n - m), (n - m) \times (n - m)$. In other words, the matrix $[T]_\beta$ is equal to $\begin{bmatrix} P & Q \\ O & R \end{bmatrix}$, where O is a zero matrix of size $(n - m) \times m$. So we have

$$\det(T) = \det[T]_\beta = \det(P)\det(R). \quad (4)$$

On the other hand, observe that P is the matrix of the restriction map $T|K$ in the basis $\{v_1, \dots, v_m\}$. So $\det(P) = \det(T|K)$. Observe also that $\bar{v}_{m+1}, \dots, \bar{v}_n$ form a basis of V/K , and that

$$\bar{T}\bar{v}_i = \sum_{j=m+1}^n r_{ij}\bar{v}_j, \quad m+1 \leq i \leq n.$$

So R is the matrix of the induced map $\bar{T} : V/K \rightarrow V/K$ in this basis. So $\det(R) = \det(\bar{T})$. Thus (4) yields the desired result. \square

We now consider the generalized eigenspaces K_λ of $A : V \rightarrow V$.

Lemma 11.19. *If λ_1, λ_2 are distinct eigenvalues of A , then $K_{\lambda_1} \cap K_{\lambda_2} = (0)$.*

Proof: Suppose the contrary, i.e. there is a nonzero vector v such that

$$(A - \lambda_1 I)^a v = 0 = (A - \lambda_2 I)^b v \quad (5)$$

for some $a, b > 0$. Let a be the smallest integer such that $(A - \lambda_1 I)^a v = 0$. Then the vector $v_1 = (A - \lambda_1 I)^{a-1} v$ is a nonzero eigenvector of A with eigenvalue λ_1 by (5). Since the subspace $K_{\lambda_1} \cap K_{\lambda_2}$ is A -invariant, and v lies in this subspace, it follows that v_1 must also lie in this subspace. Let b be the smallest integer such that $(A - \lambda_2 I)^b v_1 = 0$. Then

the vector $v_2 = (A - \lambda_2 I)^{b-1} v_1$ is a nonzero eigenvector of A with eigenvalue λ_2 . But since A commutes with $A - \lambda_2 I$, we have

$$Av_2 = A(A - \lambda_2 I)^{b-1} v_1 = (A - \lambda_2 I)^{b-1} Av_1 = \lambda_1 v_2.$$

This shows that v_2 is a *nonzero* eigenvector of A with two distinct eigenvalues λ_1, λ_2 . This is a contradiction. \square

Lemma 11.20. *Let $\lambda_1, \dots, \lambda_s$ be the distinct eigenvalues of $A : V \rightarrow V$. Then $K_{\lambda_1} + \dots + K_{\lambda_s}$ is a direct sum. That is, if $v_i \in K_{\lambda_i}$ for $i = 1, \dots, s$, and if $v_1 + \dots + v_s = 0$, then $v_i = 0$ for all i .*

Proof: Suppose the contrary, say $v_1 \neq 0$. Since $v_1 \in K_{\lambda_1}$, we have $(A - \lambda_2 I)^p v_1 \neq 0$ for all $p > 0$. For otherwise, we have $0 \neq v_1 \in K_{\lambda_1} \cap K_{\lambda_2}$, contradicting the preceding lemma. Note also that $(A - \lambda_2 I)^p v_1 \in K_{\lambda_1}$ since K_{λ_1} is A -invariant. Repeating this argument, we find that $(A - \lambda_2 I)^p \dots (A - \lambda_s I)^p v_1$ is a nonzero vector in K_{λ_1} for all $p > 0$.

On the other hand, for any given $i > 1$, we have $(A - \lambda_i I)^p v_i = 0$ for some $p > 0$ because $v_i \in K_{\lambda_i}$. So we can choose a large p so that

$$(A - \lambda_2 I)^p \dots (A - \lambda_s I)^p (v_2 + \dots + v_s) = 0.$$

But since $v_1 = -(v_2 + \dots + v_s)$, the left side is $-(A - \lambda_2 I)^p \dots (A - \lambda_s I)^p v_1 \neq 0$, a contradiction. \square

Proposition 11.21. *Let m_λ be the multiplicity of the eigenvalue λ of A , i.e. $(t - \lambda)^{m_\lambda}$ is the largest power of the linear function $t - \lambda$ appearing in the factorization of the characteristic polynomial of A . Then $\dim K_\lambda = m_\lambda$.*

Proof: For $T = A - tI : V \rightarrow V$ ($t \in \mathbf{C}$), the characteristic polynomial of A is $\det(T)$. By assumption,

$$\det(T) = (t - \lambda)^{m_\lambda} g(t)$$

for some polynomial function $g(t)$ whose factorization contains no factor $t - \lambda$.

By the second lemma above, the only eigenvalue of the restriction map $A|_{K_\lambda}$ is λ . It follows that the characteristic polynomial of this restriction map must be $(-1)^p (t - \lambda)^p$

where $p = \dim K_\lambda$. By definition this is also equal to $\det(T|K_\lambda)$. So by the first lemma above, we have

$$(t - \lambda)^{m_\lambda} g(t) = (-1)^p (t - \lambda)^p \det(\bar{T}).$$

Since $g(t)$ does not divide the function $t - \lambda$, we must have $p \leq m_\lambda$. It remains to show that $p \geq m_\lambda$.

Suppose the contrary $p < m_\lambda$. Then $\det(\bar{T})$ divides the function $t - \lambda$. But $\det(\bar{T}) = \det(\bar{A} - t\bar{I})$ is also the characteristic polynomial of $\bar{A} : V/K_\lambda \rightarrow V/K_\lambda$. This implies that λ is an eigenvalue of \bar{A} . So, for some *nonzero* $\bar{v} \in V/K_\lambda$, we have

$$\bar{A}\bar{v} = \lambda\bar{v}.$$

This says that $(A - \lambda I)v \in K_\lambda$. By definition of K_λ , this shows that

$$(A - \lambda I)^r (A - \lambda I)v = 0$$

for $r > 0$. In turn, this implies that $v \in K_\lambda$, hence \bar{v} is zero in V/K_λ , a contradiction.

This completes our proof. \square

Proof of Theorem 11.9: The preceding proposition is the first assertion of the theorem. The third lemma above shows that the subspace $K_{\lambda_1} + \cdots + K_{\lambda_s} \subset V$ is a direct sum. Hence its dimension is $m_{\lambda_1} + \cdots + m_{\lambda_s}$. But the characteristic polynomial of A is

$$(-1)^n (t - \lambda_1)^{m_{\lambda_1}} \cdots (t - \lambda_s)^{m_{\lambda_s}}$$

which must have degree $n = m_{\lambda_1} + \cdots + m_{\lambda_s}$. This shows that the subspace $K_{\lambda_1} + \cdots + K_{\lambda_s} \subset V$ has dimension n , hence must be all of V . So, we have a direct sum decomposition

$$V = K_{\lambda_1} + \cdots + K_{\lambda_s}. \quad \square$$

11.6. Appendix B

We shall prove Theorem 11.13 in this section. Throughout this section, we put $T = A - \lambda I$, where $A : K \rightarrow K$ is a fixed linear map of a complex vector space $K \neq (0)$ with a single given eigenvalue λ .

Recall that a T -cycle of size $p \geq 1$ in K is a list of vectors

$$J = \{v, Tv, \dots, T^{p-1}v\} \tag{6}$$

such that $T^{p-1}v \neq 0$ and $T^p v = 0$.

Lemma 11.22. *Let L be a T -cyclic subspace of K . Then $L \cap \text{Ker}(T)$ is one dimensional.*

Proof: By definition, L is spanned by a T -cycle (6), which contains the nonzero vector $T^{p-1}v \in \text{Ker}(T)$. So, $L \cap \text{Ker}(T)$ is at least one dimensional. Let $u \in L \cap \text{Ker}(T)$. Then u is a linear combination of J , say

$$u = a_0v + a_1Tv + \dots + a_{p-1}T^{p-1}v.$$

Applying T to this, we get

$$0 = Tu = a_0Tv + \dots + a_{p-2}T^{p-1}v.$$

By Theorem 11.12, J is linearly independent, and so $a_0 = \dots = a_{p-2} = 0$. Thus $u = a_{p-1}T^{p-1}v$. This shows that $T^{p-1}v$ is a basis of $L \cap \text{Ker}(T)$. \square

Lemma 11.23. *Let L_1, \dots, L_r be T -cyclic subspaces in K . Pick a nonzero $u_i \in L_i \cap \text{Ker}(T)$ for each i . Then u_1, \dots, u_r are linearly independent iff $L_1 + \dots + L_r$ is a direct sum.*

Proof: We show the if part first. Assume the sum is direct. Consider a linear relation $a_1u_1 + \dots + a_ru_r = 0$. Since $a_iu_i \in L_i$ for all i , we have $a_iu_i = 0$ because the sum $L_1 + \dots + L_r$ is direct. Now $a_i = 0$ because $u_i \neq 0$. This shows that u_1, \dots, u_r are linearly independent.

For the only-if part, we do induction on $\dim K$. If $\dim K = 1$, there is nothing to prove. So, suppose our assertion holds true for up to $\dim K = m - 1$. We shall prove the case $\dim K = m$. Thus, suppose the vectors $u_i \in L_i \cap \text{Ker}(T)$, $i = 1, \dots, r$ are linearly independent. We shall prove that $L_1 + \dots + L_r$ is a direct sum. First, if $\dim L_i = 1$ for all i , then $L_i = \mathbf{C}u_i$, and $L_1 + \dots + L_r$ is clearly a direct sum in this case. So, let's assume that $\dim L_1 > 1$.

Consider the projection map $K \rightarrow K/\mathbf{C}u_1$, $v \mapsto \bar{v}$. Since $\mathbf{C}u_1$ is T -invariant, T induces a map $\bar{T} : \bar{K} \rightarrow \bar{K}$, $\bar{v} \mapsto \bar{T}\bar{v}$. Now, suppose L is a nonzero T -cyclic subspace of K . We claim that *the image \bar{L} under the projection map is a \bar{T} -cyclic subspace of \bar{K}* . First, L is spanned by a T -cycle, say (6). Then \bar{L} is spanned by

$$\bar{J} = \{\bar{v}, \bar{T}\bar{v}, \dots, \bar{T}^{p-1}\bar{v}\}.$$

If $\bar{v} = 0$, then $\bar{L} = (0)$, which is trivially \bar{T} -cyclic. If $\bar{v} \neq 0$, then there is a smallest positive integer $d \leq p$ such that $\bar{T}^d \bar{v} = 0$. In this case, \bar{L} is spanned by $\{\bar{v}, \bar{T}\bar{v}, \dots, \bar{T}^{d-1}\bar{v}\} \subset \bar{J}$, which is a \bar{T} -cycle in \bar{K} . In any case, \bar{L} is \bar{T} -cyclic. So, the image \bar{L}_i of each L_i is a \bar{T} -cyclic subspace of \bar{K} . Note that $\bar{u}_2 \in \bar{L}_2$ is a nonzero vector lying in $\text{Ker}(\bar{T})$, because $u_2 \notin \mathbf{C}u_1$ by linear independence and because $\bar{T}\bar{u}_2 = \overline{T u_2} = 0$. Likewise for $i = 2, \dots, r$, we have

$$0 \neq \bar{u}_i \in \bar{L}_i \cap \text{Ker}(\bar{T}).$$

Also $\bar{L}_1 \neq (0)$ since $L_1 \not\subset \mathbf{C}u_1$, because $\dim L_1 > 1$. By the preceding lemma, \bar{L}_1 contains a nonzero vector $\bar{w} \in \text{Ker}(\bar{T})$.

We claim that $\bar{w}, \bar{u}_2, \dots, \bar{u}_r$ are linearly independent. Consider a linear relation $a_1 \bar{w} + a_2 \bar{u}_2 + \dots + a_r \bar{u}_r = 0$ in \bar{K} . So, in K , we have

$$aw + a_2 u_2 + \dots + a_r u_r = a_1 u_1 \quad (7)$$

for some $a_1 \in \mathbf{C}$. Applying T to this, we get $aTw = 0$ because $u_i \in \text{Ker}(T)$. So either $Tw = 0$ or $a = 0$. Since $\bar{w} \neq 0$, it follows that $w \notin \mathbf{C}u_1$. But $L_1 \cap \text{Ker}(T) = \mathbf{C}u_1$ by preceding lemma. So, $w \notin \text{Ker}(T)$, i.e. $Tw \neq 0$. Hence $a = 0$. Now (7) implies that $a_i = 0$ for all i , because the u_i are linearly independent.

So, we have shown that $\bar{w}, \bar{u}_2, \dots, \bar{u}_r \in \text{Ker}(T)$, lying respectively in the \bar{T} -cyclic subspaces $\bar{L}_1, \dots, \bar{L}_r$, are linearly independent vectors in \bar{K} . Since $\dim \bar{K} = m - 1$, our inductive hypothesis implies that

$$\bar{L}_1 + \dots + \bar{L}_r \quad (8)$$

is a direct sum in \bar{K} . We now show that $L_1 + \dots + L_r$ is a direct sum in K . Let $v_i \in L_i$ such that $v_1 + \dots + v_r = 0$. Applying the projection map to this, we get $\bar{v}_1 + \dots + \bar{v}_r = 0$. Since (8) is a direct sum and $\bar{v}_i \in \bar{L}_i$, we have $\bar{v}_i = 0$ for all i . In particular for $i = 2$,

$$v_2 = au_1$$

for some $a \in \mathbf{C}$. Since the right side lies in $\text{Ker}(T)$ and the left side lies in L_2 , both sides lie in $L_2 \cap \text{Ker}(T) = \mathbf{C}u_2$. Thus, we have $v_2 = au_1 \in \mathbf{C}u_2$, hence $a = 0$ and so $v_2 = 0$, because u_1, u_2 are linearly independent. Likewise, $v_i = 0$ for $i = 2, \dots, r$. Finally, $v_1 + \dots + v_r = 0$ implies that $v_1 = 0$ as well. This proves that $L_1 + \dots + L_r$ is a direct sum when $\dim K = m$. \square

To prove Theorem 11.13, the preceding lemma suggests that starting from linearly independent eigenvectors $u_1, \dots, u_r \in T$, we should find T -cyclic subspaces $L_i \ni u_i$. Then

hope to get $K = L_1 + \cdots + L_r$. Unfortunately, doing it this way, the T -cyclic subspaces may fall short of saturating K , i.e. we may end up with just $K \supsetneq L_1 + \cdots + L_r$. Here is an example.

Example 11.24.

Let $T : \mathbf{C}^4 \rightarrow \mathbf{C}^4$ be the linear map defined by

$$e_1 \mapsto 0, \quad e_2 \mapsto e_1, \quad e_3 \mapsto e_2, \quad e_4 \mapsto 0$$

where the e_i form the standard basis of \mathbf{C}^4 . Note that in this basis, the matrix of T is already a Jordan matrix with a single eigenvalue 0 and two Jordan blocks of sizes 3,1. (Write it down!) The eigenspace is

$$\text{Ker}(T) = \mathbf{C}e_1 + \mathbf{C}e_4.$$

Note that $J_1 = \{e_2 + e_4, e_1\}$ and $J_2 = \{e_4\}$ are T -cycles. The T -cyclic subspaces L_1, L_2 they each spans contain the eigenvectors e_1, e_4 respectively. There are no more independent eigenvectors we can use to make more T -cycle. Yet, the T -cyclic subspaces we get from J_1, J_2 add up to just 3 dimension, not enough to give $K = \mathbf{C}^4$. Note that we cannot “extend” either J_1, J_2 , as it stands, by adding more vectors to the left, because $e_2 + e_4, e_4$ are not in $\text{Im}(T)$. Roughly, the problem is that J_1 is not a T -cycle that is as large as while containing the eigenvector e_1 . In fact, here is one that is as big as it can be: $\{e_3, e_2, e_1\}$. Note that this new T -cycle together with $\{e_4\}$ are enough to give T -cyclic subspaces that add up to 4 dimension. Note also that the two eigenvectors $e_1 \pm e_4$ cannot be extended at all to a larger T -cycle.

Proof of Theorem 11.13: We shall do induction on $\dim K$. Again for $\dim K = 1$, there is nothing to prove. Suppose the theorem holds for up to $\dim K = m - 1$. We shall prove the case $\dim K = m$. Since $\text{Ker}(T) \neq 0$, $\dim TK < \dim K$ by the dimension relation. Since $TK = \text{Im}(T)$ is T -invariant, we can apply our inductive hypothesis to the restriction map $T|_{TK}$, which we still call T . Thus there exists T -cyclic subspaces M_1, \dots, M_s of TK , and a direct sum decomposition

$$TK = M_1 + \cdots + M_s. \tag{9}$$

Consider one of the summands, say M_1 . Since M_1 is T -cyclic, there exists $w \in TK$, say $w = Tv$, such that M_1 is spanned by a T -cycle $\{Tv, \dots, T^{d-1}v\}$ in TK . Adjoining v to

this T -cycle, we get a new T -cycle of size d in K . Let L_1 be the T -cyclic subspace of K spanned by the new T -cycle. By Theorem 11.12, $\dim L_1 = \dim M_1 + 1$. (Note that the space L_1 , though not $\dim L$, depends on the choice of v .) Repeat this construction for each M_i and get L_i with

$$\dim L_i = \dim M_i + 1.$$

Let u_1, \dots, u_s be nonzero vectors in the respective one dimensional spaces $L_i \cap \text{Ker}(T)$ (cf. Lemma 11.22.) Since (9) is a direct sum, these vectors are linearly independent. So, we can extend $\{u_1, \dots, u_s\} \subset \text{Ker}(T)$ to a basis $\{u_1, \dots, u_s, \dots, u_r\}$ of $\text{Ker}(T)$. We can regard $\{u_i\}$, for $i = s + 1, \dots, r$, as T -cycles in K , and let L_i be the one dimensional T -cyclic subspace $\mathbf{C}u_i$, for $i = s + 1, \dots, r$. Thus we now have r T -cyclic subspaces $L_i \ni u_i$ in K and linearly independent eigenvectors u_1, \dots, u_r of T . By Lemma 11.23,

$$L_1 + \dots + L_r \subset K$$

is a direct sum. It remains to show that this direct sum has dimension $\dim K$. Its dimension is

$$\sum_{i=1}^s L_i + \sum_{i=s+1}^r L_i = \sum_{i=1}^s (\dim M_i + 1) + (r-s) = \dim TK + s + (r-s) = \dim TK + \dim \text{Ker}(T)$$

But the last expression is equal to $\dim K$ by the dimension relation. \square

11.7. Homework

7. Prove Schwarz's inequality for \mathbf{C}^n by imitating the case of \mathbf{R}^n .
8. An $n \times n$ matrix A is said to be unitary if $AA^\dagger = I$. This is the complex analogue of the notion of an orthogonal matrix. Show that the product of two unitary matrices is unitary and the inverse of a unitary matrix is also unitary. Prove that $n \times n$ unitary matrices are in 1-1 correspondence with (ordered) orthonormal bases of \mathbf{C}^n .
9. * By imitating the proof of the spectral theorem for real symmetric matrices, prove that if A is a hermitian matrix, there is a unitary matrix B such that BAB^\dagger is diagonal.

10. Suppose that a linear map $A : V \rightarrow V$ has characteristic polynomial $(t - \lambda_1)^3(t - \lambda_2)^3$ where $\lambda_1 \neq \lambda_2$, and that its generalized eigenspaces are

$$K_{\lambda_1} = \text{Ker}(A - \lambda_1 I)^3 \supsetneq \text{Ker}(A - \lambda_1 I)^2, \quad K_{\lambda_2} = \text{Ker}(A - \lambda_2 I)^3 \supsetneq \text{Ker}(A - \lambda_2 I).$$

Find a Jordan matrix of A .

11. Let $f(t) = a_m t^m + \cdots + a_1 t + a_0$ be a polynomial function. If A is an $n \times n$ matrix, we define $f(A)$ to be the matrix

$$f(A) = a_m A^m + \cdots + a_1 A + a_0 I.$$

Show that if C is any invertible matrix, then $f(CAC^{-1}) = Cf(A)C^{-1}$. In particular, if $f(A) = 0$ then $f(CAC^{-1}) = 0$.

12. Show that if A, B are similar matrices, then they have the same characteristic polynomial. (Hint: $\det(PQ) = \det(P)\det(Q)$.)

13. Let B be a Jordan matrix with the $k_i \times k_i$ blocks B_i along the diagonal as given in Definition 11.6. Let e_1, \dots, e_n be the standard unit (column) vectors in \mathbf{C}^n . Show that for each $i = 1, \dots, p$, we have

$$(B - \lambda_i I)^{k_i} e_j = 0, \quad k_{i-1} + 1 \leq j \leq k_i, \quad (k_0 = 0.)$$

Use this to show that

$$(B - \lambda_1 I)^{k_1} \cdots (B - \lambda_p I)^{k_p} = 0.$$

Now argue that if B is a Jordan canonical form of A , then

$$(A - \lambda_1 I)^{k_1} \cdots (A - \lambda_p I)^{k_p} = 0.$$

14. (continuing the preceding problem) Let $f(t)$ be the characteristic polynomial $\det(A - tI)$ of A . Prove that

$$f(t) = (-1)^n (t - \lambda_1)^{k_1} \cdots (t - \lambda_p)^{k_p}.$$

(Hint: Find the characteristic polynomial of B first.) Finally, conclude that $f(A) = 0$. This result is known as the Cayley-Hamilton theorem.

15. *Minimal polynomial.* Let A be a given nonzero $n \times n$ matrix.
- Prove that there is a lowest degree nonconstant polynomial function $g(t)$ such that $g(A) = 0$. (Hint: What is the dimension of the space of $n \times n$ matrices $M(n, n)$?)
 - Prove that any two such polynomials must be a scalar multiple of one another. Thus there is a unique such polynomial $g(t)$ if we further assume that its leading term has the form t^m . This unique polynomial, determined by A , is called the minimal polynomial of A .

Note: It can be shown that if $f(t)$ is any polynomial function such that $f(A) = 0$, then $f(t)$ is divisible by the minimal polynomial $g(t)$, i.e. $f(t) = g(t)h(t)$ for some polynomial function $h(t)$. In particular if $f(t)$ is the characteristic polynomial of A , the Cayley-Hamilton theorem implies that $f(t)$ is divisible by $g(t)$.

In the problems below, let $T : K \rightarrow K$ be a linear map of a finite dimensional vector space K , and assume that there is a smallest positive integer q such that $K = \text{Ker}(T^q)$.

16. Show that

$$K \supseteq TK \supseteq T^2K \supseteq \cdots \supseteq T^qK = (0).$$

A nonzero vector $u \in K$ is said to have depth d if d is the largest positive integer such that $u \in T^{d-1}K$. If u is a nonzero vector in $\text{Ker}(T)$ and has depth d , show that there is a T -cyclic subspace L of K of dimension d such that $u \in L$.

17. (cf. Appendix B.) Show that if L is a T -cyclic subspace of K , then $L \cap \text{Ker}(T)$ is one dimensional.
18. A T -cyclic subspace L of K is called *maximal* if $\dim L = \text{depth}(u)$ for a nonzero vector u in $L \cap \text{Ker}(T)$. Show that every nonzero vector u in $\text{Ker}(T)$ is contained in a maximal T -cyclic subspace of K .
19. Let L_1, \dots, L_r be T -cyclic subspaces of K such that $L_1 + \cdots + L_r$ is a *direct sum*. If $K = L_1 + \cdots + L_r$, show that each L_i is maximal. (Hint: Assume L_1 is not

maximal. Use the preceding problem to construct another direct sum bigger than K .)

20. (cf. Appendix A.) Let L be a T -invariant subspace of K . Consider the quotient space $\bar{K} = K/L$ and the projection map $K \rightarrow \bar{K}$, $v \mapsto \bar{v}$. Show that the map $\bar{T} : \bar{K} \rightarrow \bar{K}$, $v \mapsto \overline{Tv}$ is well-defined, and that \bar{T}^q coincides with the zero map.
21. Let u be a nonzero vector with $\text{depth}(u) = q$. Show that $u \in \text{Ker}(T)$, hence $L = \mathbf{C}u$ is T -invariant. Consider the quotient space $\bar{K} = K/L$ and the map $\bar{T} : \bar{K} \rightarrow \bar{K}$. Show that if w is a nonzero vector in $\text{Ker}(T)$ and \bar{w} is nonzero, then $\text{depth}(\bar{w}) = \text{depth}(w)$.
22. * Let T be the set of all diagonal invertible $n \times n$ complex matrices, and \mathbf{C}^\times be the set of nonzero complex numbers. Let $f : T \rightarrow \mathbf{C}^\times$ be a continuous map such that $f(AB) = f(A)f(B)$ and $f(A^{-1}) = f(A)^{-1}$ for all $A, B \in T$. For continuity, f is thought of as a function of n variables (the diagonal entries.) Prove that $f(A) = \det(A)^k$ for some integer k . (Hint: First consider the case $n = 1$, so that $T = \mathbf{C}^\times$. Restrict the map f to the unit circle $S \subset T$, consisting of complex numbers a with $|a| = 1$. Show that $f : S \rightarrow \mathbf{C}^\times$ must be of the form $f(a) = a^k$.)
23. * Let G be the set of all invertible $n \times n$ complex matrices and $f : G \rightarrow \mathbf{C}^\times$ be a continuous map such that $f(AB) = f(A)f(B)$ and $f(A^{-1}) = f(A)^{-1}$ for all $A, B \in G$. Prove that $f(A) = \det(A)^k$ for some integer k . (Hint: Do the preceding problem first. Then restrict f to $T \subset G$. Observe that diagonalizable elements of G form a dense subset of G and that $f(A) = \det(A)^k$ holds for diagonalizable A .)

