Chapter 2
Review of Linear Algebra

This chapter reviews the linear algebra that we shall assume throughout the book. Proofs of standard results are mostly omitted. The reader can consult a linear algebra text such as [4] for details. In this book all vector spaces considered will be finite dimensional over the field $\mathbb{C}$ of complex numbers.

2.1 Basic Definitions and Notation

This section introduces some basic notions from linear algebra. We start with some notation, not all of which belongs to linear algebra. Let V and W be vector spaces.

- If X is a set of vectors, then $\mathbb{C}X = \operatorname{Span} X$.
- $M_{mn}(\mathbb{C}) = \{m \times n \text{ matrices with entries in } \mathbb{C}\}$.
- $M_n(\mathbb{C}) = M_{nn}(\mathbb{C})$.
- $\operatorname{Hom}(V, W) = \{A\colon V \longrightarrow W \mid A \text{ is a linear map}\}$.
- $\operatorname{End}(V) = \operatorname{Hom}(V, V)$ (the *endomorphism ring* of V).
- $GL(V) = \{A \in \operatorname{End}(V) \mid A \text{ is invertible}\}$ (known as the *general linear group of* V).
- $GL_n(\mathbb{C}) = \{A \in M_n(\mathbb{C}) \mid A \text{ is invertible}\}$.
- The identity matrix/linear transformation is denoted I, or I_n if we wish to emphasize the dimension n.
- $\mathbb{Z}$ is the ring of integers.
- $\mathbb{N}$ is the set of non-negative integers.
- $\mathbb{Q}$ is the field of rational numbers.
- $\mathbb{R}$ is the field of real numbers.
- $\mathbb{Z}/n\mathbb{Z} = \{[0], \ldots, [n-1]\}$ is the ring of integers modulo n.
- R^* denotes the group of units (i.e., invertible elements) of a ring R.
- S_n is the group of permutations of $\{1, \ldots, n\}$, i.e., the *symmetric group* on n letters.
- The identity permutation is denoted Id.

B. Steinberg, *Representation Theory of Finite Groups: An Introductory Approach*, Universitext, DOI 10.1007/978-1-4614-0776-8_2,

Elements of $\mathbb{C}^n$ will be written as n-tuples or as column vectors, as is convenient.

If $A \in M_{mn}(\mathbb{C})$, we sometimes write A_{ij} for the entry in row i and column j. We may also write $A = (a_{ij})$ to mean the matrix with a_{ij} in row i and column j. If k, ℓ, m, and n are natural numbers, then matrices in $M_{mk,\ell n}(\mathbb{C})$ can be viewed as $m \times n$ block matrices with blocks in $M_{k\ell}(\mathbb{C})$. If we view an $mk \times \ell n$ matrix A as a block matrix, then we write $[A]_{ij}$ for the $k \times \ell$ matrix in the i, j block, for $1 \leq i \leq m$ and $1 \leq j \leq n$.

Definition 2.1.1 (Coordinate vector). If V is a vector space with basis $B = \{b_1, \ldots, b_n\}$ and $v = c_1 b_1 + \cdots + c_n b_n$ is a vector in V, then the *coordinate vector* of v with respect to the basis B is the vector $[v]_B = (c_1, \ldots, c_n) \in \mathbb{C}^n$. The map $T\colon V \longrightarrow \mathbb{C}^n$ given by $Tv = [v]_B$ is a vector space isomorphism that we sometimes call *taking coordinates* with respect to B.

Suppose that $T\colon V \longrightarrow W$ is a linear transformation and B, B' are bases for V, W, respectively. Let $B = \{v_1, \ldots, v_n\}$ and $B' = \{w_1, \ldots, w_m\}$. Then the *matrix of* T with respect to the bases B, B' is the $m \times n$ matrix $[T]_{B,B'}$ whose jth column is $[Tv_j]_{B'}$. In other words, if

$$Tv_j = \sum_{i=1}^{m} a_{ij} w_i,$$

then $[T]_{B,B'} = (a_{ij})$. When $V = W$ and $B = B'$, then we write simply $[T]_B$ for $[T]_{B,B}$.

The *standard basis* for $\mathbb{C}^n$ is the set $\{e_1, \ldots, e_n\}$ where e_i is the vector with 1 in the ith coordinate and 0 in all other coordinates. So when $n = 3$, we have

$$e_1 = (1, 0, 0), \quad e_2 = (0, 1, 0), \quad e_3 = (0, 0, 1).$$

Throughout we will abuse the distinction between $\mathrm{End}(\mathbb{C}^n)$ and $M_n(\mathbb{C})$ and the distinction between $GL(\mathbb{C}^n)$ and $GL_n(\mathbb{C})$ by identifying a linear transformation with its matrix with respect to the standard basis.

Suppose $\dim V = n$ and $\dim W = m$. Then by choosing bases for V and W and sending a linear transformation to its matrix with respect to these bases we see that:

$$\begin{aligned} \mathrm{End}(V) &\cong M_n(\mathbb{C}); \\ GL(V) &\cong GL_n(\mathbb{C}); \\ \mathrm{Hom}(V, W) &\cong M_{mn}(\mathbb{C}). \end{aligned}$$

Notice that $GL_1(\mathbb{C}) \cong \mathbb{C}^*$ and so we shall always work with the latter. We indicate W is a subspace of V by writing $W \leq V$.

If $W_1, W_2 \leq V$, then by definition

$$W_1 + W_2 = \{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\}.$$

This is the smallest subspace of V containing W_1 and W_2. If, in addition, $W_1 \cap W_2 = \{0\}$, then $W_1 + W_2$ is called a *direct sum*, written $W_1 \oplus W_2$. As vector spaces, $W_1 \oplus W_2 \cong W_1 \times W_2$ via the map $W_1 \times W_2 \longrightarrow W_1 \oplus W_2$ given by $(w_1, w_2) \mapsto w_1 + w_2$. In fact, if V and W are any two vector spaces, one can form their *external direct sum* by setting $V \oplus W = V \times W$. Note that

$$\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2.$$

More precisely, if B_1 is a basis for W_1 and B_2 is a basis for W_2, then $B_1 \cup B_2$ is a basis for $W_1 \oplus W_2$.

2.2 Complex Inner Product Spaces

Recall that if $z = a + bi \in \mathbb{C}$, then its *complex conjugate* is $\overline{z} = a - bi$. In particular, $z\overline{z} = a^2 + b^2 = |z|^2$. An *inner product* on V is a map

$$\langle \cdot, \cdot \rangle \colon V \times V \longrightarrow \mathbb{C}$$

such that, for $v, w, v_1, v_2 \in V$ and $c_1, c_2 \in \mathbb{C}$:

- $\langle c_1 v_1 + c_2 v_2, w \rangle = c_1 \langle v_1, w \rangle + c_2 \langle v_2, w \rangle$;
- $\langle w, v \rangle = \overline{\langle v, w \rangle}$;
- $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ if and only if $v = 0$.

A vector space equipped with an inner product is called an *inner product space*. The *norm* $\|v\|$ of a vector v in an inner product space is defined by $\|v\| = \sqrt{\langle v, v \rangle}$.

Example 2.2.1. The *standard inner product* on $\mathbb{C}^n$ is given by

$$\langle (a_1, \ldots, a_n), (b_1, \ldots, b_n) \rangle = \sum_{i=1}^{n} a_i \overline{b_i}.$$

Two important properties of inner products are the *Cauchy–Schwarz inequality*

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

and the *triangle inequality*

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

Recall that two vectors v, w in an inner product space V are said to be *orthogonal* if $\langle v, w \rangle = 0$. A subset of V is called *orthogonal* if its elements are pairwise orthogonal. If, in addition, the norm of each vector is 1, the set is termed *orthonormal*. An orthogonal set of non-zero vectors is linearly independent. In particular, any orthonormal set is linearly independent.

Every inner product space has an orthonormal basis. One can obtain an orthonormal basis from an arbitrary basis using the Gram–Schmidt process [4, Theorem 15.9]. If $B = \{e_1, \ldots, e_n\}$ is an orthonormal basis for an inner product space V and $v \in V$, then

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

In other words,

$$[v]_B = (\langle v, e_1\rangle, \ldots, \langle v, e_n\rangle).$$

Example 2.2.2. For a finite set X, the set $\mathbb{C}^X = \{f \colon X \longrightarrow \mathbb{C}\}$ is a vector space with pointwise operations. Namely, one defines

$$(f + g)(x) = f(x) + g(x);$$
$$(cf)(x) = cf(x).$$

For each $x \in X$, define a function $\delta_x \colon X \longrightarrow \mathbb{C}$ by

$$\delta_x(y) = \begin{cases} 1 & x = y \\ 0 & x \neq y. \end{cases}$$

There is a natural inner product on $\mathbb{C}^X$ given by

$$\langle f, g\rangle = \sum_{x \in X} f(x)\overline{g(x)}.$$

The set $\{\delta_x \mid x \in X\}$ is an orthonormal basis with respect to this inner product. If $f \in \mathbb{C}^X$, then its unique expression as a linear combination of the δ_x is given by

$$f = \sum_{x \in X} f(x)\delta_x.$$

Consequently, $\dim \mathbb{C}^X = |X|$.

Direct sum decompositions are easy to obtain in inner product spaces. If $W \leq V$, then the *orthogonal complement* of W is the subspace

$$W^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ for all } w \in W\}.$$

Proposition 2.2.3. *Let V be an inner product space and $W \leq V$. Then there is a direct sum decomposition $V = W \oplus W^\perp$.*

Proof. First, if $w \in W \cap W^\perp$ then $\langle w, w\rangle = 0$ implies $w = 0$; so $W \cap W^\perp = \{0\}$. Let $v \in V$ and suppose that $\{e_1, \ldots, e_m\}$ is an orthonormal basis for W. Put $\hat{v} = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$ and $z = v - \hat{v}$. Then $\hat{v} \in W$. We claim that $z \in W^\perp$.

To prove this, it suffices to show $\langle z, e_i\rangle = 0$ for all $i = 1, \ldots, m$. To this effect we compute

$$\langle z, e_i\rangle = \langle v, e_i\rangle - \langle \hat{v}, e_i\rangle = \langle v, e_i\rangle - \langle v, e_i\rangle = 0$$

because $\{e_1, \ldots, e_m\}$ is an orthonormal set. As $v = \hat{v} + z$, it follows that $V = W + W^{\perp}$. This completes the proof. □

We continue to assume that V is an inner product space.

Definition 2.2.4 (Unitary operator). A linear operator $U \in GL(V)$ is said to be *unitary* if

$$\langle Uv, Uw\rangle = \langle v, w\rangle$$

for all $v, w \in V$.

Notice that if U is unitary and $v \in \ker U$, then $0 = \langle Uv, Uv\rangle = \langle v, v\rangle$ and so $v = 0$. Thus unitary operators are invertible. The set $U(V)$ of unitary maps is a subgroup of $GL(V)$.

If $A = (a_{ij}) \in M_{mn}(\mathbb{C})$ is a matrix, then its *transpose* is the matrix $A^T = (a_{ji}) \in M_{nm}(\mathbb{C})$. The *conjugate* of A is $\overline{A} = (\overline{a_{ij}})$. The *conjugate-transpose* or *adjoint* of A is the matrix $A^* = \overline{A^T}$. One can verify directly the equality $(AB)^* = B^*A^*$. Routine computation shows that if $v \in \mathbb{C}^n$ and $w \in \mathbb{C}^m$, then

$$\langle Av, w\rangle = \langle v, A^*w\rangle \tag{2.1}$$

where we use the standard inner product on $\mathbb{C}^m$ and $\mathbb{C}^n$. Indeed, viewing vectors as column vectors one has $\langle v_1, v_2\rangle = \overline{v_1^* v_2}$ and so $\langle Av, w\rangle = \overline{(Av)^* w} = \overline{v^*(A^*w)} = \langle v, A^*w\rangle$.

With respect to the standard inner product on $\mathbb{C}^n$, the linear transformation associated to a matrix $A \in GL_n(\mathbb{C})$ is unitary if and only if $A^{-1} = A^*$ [4, Theorem 32.7]; such a matrix is thus called *unitary*. We denote by $U_n(\mathbb{C})$ the group of all $n \times n$ unitary matrices. A matrix $A \in M_n(\mathbb{C})$ is called *self-adjoint* if $A^* = A$. A matrix A is *symmetric* if $A^T = A$. If A has real entries, then A is self-adjoint if and only if A is symmetric.

More generally, if T is a linear operator on an inner product space V, then $T^*\colon V \longrightarrow V$ is the unique linear operator satisfying $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all $v, w \in V$. It is called the *adjoint* of T. If B is an orthonormal basis for V, then $[T^*]_B = [T]_B^*$. The operator T is *self-adjoint* if $T = T^*$, or equivalently if the matrix of T with respect to some (equals any) orthonormal basis of V is self-adjoint.

2.3 Further Notions from Linear Algebra

If $X \subseteq \mathrm{End}(V)$ and $W \leq V$, then W is called *X-invariant* if, for any $A \in X$ and any $w \in W$, one has $Aw \in W$, i.e., $XW \subseteq W$.

A key example comes from the theory of eigenvalues and eigenvectors. Recall that $\lambda \in \mathbb{C}$ is an *eigenvalue* of $A \in \mathrm{End}(V)$ if $\lambda I - A$ is not invertible, or in other words, if $Av = \lambda v$ for some $v \neq 0$. The *eigenspace* corresponding to λ is the set

$$V_\lambda = \{v \in V \mid Av = \lambda v\},$$

which is a subspace of V. Note that if $v \in V_\lambda$, then $A(Av) = A(\lambda v) = \lambda Av$, so $Av \in V_\lambda$. Thus V_λ is A-invariant. On the other hand, if $W \leq V$ is A-invariant with $\dim W = 1$ (that is, W is a line), then $W \subseteq V_\lambda$ for some λ. In fact, if $w \in W \setminus \{0\}$, then $\{w\}$ is a basis for W. Since $Aw \in W$, we have that $Aw = \lambda w$ for some $\lambda \in \mathbb{C}$. So w is an eigenvector with eigenvalue λ, whence $w \in V_\lambda$. Thus $W \subseteq V_\lambda$.

The *trace* of a matrix $A = (a_{ij})$ is defined by

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Some basic facts concerning the trace function $\mathrm{Tr}\colon M_n(\mathbb{C}) \longrightarrow \mathbb{C}$ are that Tr is linear and $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. Consequently, $\mathrm{Tr}(PAP^{-1}) = \mathrm{Tr}(P^{-1}PA) = \mathrm{Tr}(A)$ for any invertible matrix P. In particular, if $T \in \mathrm{End}(V)$, then $\mathrm{Tr}(T)$ makes sense: choose any basis for the vector space V and compute the trace of the associated matrix.

The *determinant* $\det A$ of a matrix is defined as follows:

$$\det A = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \cdot a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

We recall that

$$\mathrm{sgn}(\sigma) = \begin{cases} 1 & \sigma \text{ is even} \\ -1 & \sigma \text{ is odd.} \end{cases}$$

The key properties of the determinant that we shall use are:

- $\det A \neq 0$ if and only if $A \in GL_n(\mathbb{C})$;
- $\det(AB) = \det A \cdot \det B$;
- $\det(A^{-1}) = (\det A)^{-1}$.

In particular, one has $\det(PAP^{-1}) = \det A$ and so we can define, for any $T \in \mathrm{End}(V)$, the determinant by choosing a basis for V and computing the determinant of the corresponding matrix for T.

The *characteristic polynomial* $p_A(x)$ of a linear operator A on an n-dimensional vector space V is given by $p_A(x) = \det(xI - A)$. This is a monic polynomial of degree n (i.e., has leading coefficient 1) and the roots of $p_A(x)$ are precisely the eigenvalues of A. The classical Cayley–Hamilton theorem says that any linear operator is a zero of its characteristic polynomial [4, Corollary 24.7].

Theorem 2.3.1 (Cayley–Hamilton). *Let $p_A(x)$ be the characteristic polynomial of A. Then $p_A(A) = 0$.*

If $A \in \mathrm{End}(V)$, *the minimal polynomial* of A, denoted $m_A(x)$, is the smallest degree monic polynomial $f(x)$ such that $f(A) = 0$.

Proposition 2.3.2. *If $q(A) = 0$ then $m_A(x) \mid q(x)$.*

Proof. Write $q(x) = m_A(x)f(x) + r(x)$ with either $r(x) = 0$, or $\deg(r(x)) < \deg(m_A(x))$. Then

$$0 = q(A) = m_A(A)f(A) + r(A) = r(A).$$

By minimality of $m_A(x)$, we conclude that $r(x) = 0$. □

Corollary 2.3.3. *If $p_A(x)$ is the characteristic polynomial of A, then $m_A(x)$ divides $p_A(x)$.*

The relevance of the minimal polynomial is that it provides a criterion for diagonalizability of a matrix, among other things. A standard result from linear algebra is the following characterization of diagonalizable matrices, cf. [4, Theorem 23.11].

Theorem 2.3.4. *A matrix $A \in M_n(\mathbb{C})$ is diagonalizable if and only if $m_A(x)$ has no repeated roots.*

Example 2.3.5. The diagonal matrix

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

has $m_A(x) = (x-1)(x-3)$, whereas $p_A(x) = (x-1)^2(x-3)$. On the other hand, the matrix

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

has $m_B(x) = (x-1)^2 = p_B(x)$ and so is not diagonalizable.

One of the main results from linear algebra is the spectral theorem for self-adjoint matrices. We sketch a proof since it is indicative of several proofs later in the text.

Theorem 2.3.6 (Spectral Theorem). *Let $A \in M_n(\mathbb{C})$ be self-adjoint. Then there is a unitary matrix $U \in U_n(\mathbb{C})$ such that U^*AU is diagonal. Moreover, the eigenvalues of A are real.*

Proof. First we verify that the eigenvalues are real. Let λ be an eigenvalue of A with corresponding eigenvector v. Then

$$\lambda\langle v, v\rangle = \langle Av, v\rangle = \langle v, A^*v\rangle = \langle v, Av\rangle = \overline{\langle Av, v\rangle} = \overline{\lambda}\langle v, v\rangle$$

and hence $\lambda = \overline{\lambda}$ because $\langle v, v\rangle > 0$. Thus λ is real.

To prove the remainder of the theorem, it suffices to show that $\mathbb{C}^n$ has an orthonormal basis B of eigenvectors for A. One can then take U to be a matrix whose columns are the elements of B. We proceed by induction on n, the case $n = 1$ being trivial. Assume the theorem is true for all dimensions smaller than n. Let λ be an eigenvalue of A with corresponding eigenspace V_λ. If $V_\lambda = \mathbb{C}^n$, then A already is diagonal and there is nothing to prove. So we may assume that V_λ is a proper subspace; it is of course non-zero. Then $\mathbb{C}^n = V_\lambda \oplus V_\lambda^\perp$ by Proposition 2.2.3. We claim that $V_\lambda^\perp$ is A-invariant. Indeed, if $v \in V_\lambda$ and $w \in V_\lambda^\perp$, then

$$\langle Aw, v\rangle = \langle w, A^* v\rangle = \langle w, Av\rangle = \langle w, \lambda v\rangle = 0$$

and so $Aw \in V_\lambda^\perp$. Note that $V_\lambda^\perp$ is an inner product space in its own right by restricting the inner product on V, and moreover the restriction of A to $V_\lambda^\perp$ is still self-adjoint. Since $\dim V_\lambda^\perp < n$, an application of the induction hypothesis yields that $V_\lambda^\perp$ has an orthonormal basis B' of eigenvectors for A. Let B be any orthonormal basis for V_λ. Then $B \cup B'$ is an orthonormal basis for $\mathbb{C}^n$ consisting of eigenvectors for A, as required. □

Exercises

Exercise 2.1. Suppose that $A, B \in M_n(\mathbb{C})$ are commuting matrices, i.e., $AB = BA$. Let V_λ be an eigenspace of A. Show that V_λ is B-invariant.

Exercise 2.2. Let V be an n-dimensional vector space and B a basis. Prove that the map $F\colon \operatorname{End}(V) \longrightarrow M_n(\mathbb{C})$ given by $F(T) = [T]_B$ is an isomorphism of unital rings.

Exercise 2.3. Let V be an inner product space and let $W \leq V$ be a subspace. Let $v \in V$ and define $\hat{v} \in W$ as in the proof of Proposition 2.2.3. Prove that if $w \in W$ with $w \neq \hat{v}$, then $\|v - \hat{v}\| < \|v - w\|$. Deduce that $\hat{v}$ is independent of the choice of orthonormal basis for W. It is called the *orthogonal projection* of v onto W.

Exercise 2.4. Prove that $(AB)^* = B^* A^*$.

Exercise 2.5. Prove that $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$.

Exercise 2.6. Let V be an inner product space and let $T\colon V \longrightarrow V$ be a linear transformation. Show that T is self-adjoint if and only if V has an orthonormal basis of eigenvectors of T and the eigenvalues of T are real. (Hint: one direction is a consequence of the spectral theorem.)

Exercise 2.7. Let V be an inner product space. Show that $U \in \operatorname{End}(V)$ is unitary if and only if $\|Uv\| = \|v\|$ for all vectors $v \in V$. (Hint: use the polarization formula $\langle v, w\rangle = 1/4\left[\|v + w\|^2 - \|v - w\|^2\right]$.)

Exercise 2.8. Prove that if $A \in M_n(\mathbb{C})$, then there is an upper triangular matrix T and an invertible matrix P such that $P^{-1}AP = T$. (Hint: use induction on n. Look at the proof of the spectral theorem for inspiration.)

Exercise 2.9. This exercise sketches a proof of the Cayley–Hamilton Theorem using a little bit of analysis.

1. Use Exercise 2.8 to reduce to the case when A is an upper triangular matrix.
2. Prove the Cayley–Hamilton theorem for diagonalizable operators.
3. Identifying $M_n(\mathbb{C})$ with $\mathbb{C}^{n^2}$, show that the mapping $M_n(\mathbb{C}) \longrightarrow M_n(\mathbb{C})$ given by $A \mapsto p_A(A)$ is continuous. (Hint: the coefficients of $p_A(x)$ are polynomials in the entries of A.)
4. Prove that every upper triangular matrix is a limit of matrices with distinct eigenvalues (and hence diagonalizable).
5. Deduce the Cayley–Hamilton theorem.