Chapter 4
Spectral Theory

In the previous chapter, we studied spacial operators: the self-adjoint operator and normal operators. In this chapter, we study a general linear map \( A \) that maps a finite dimensional vector space \( V \) over a field \( F \) into itself. Since for a general mapping, a Euclidean structure has little use, we shall ignore it. Instead, we shall assume that the field is closed; i.e., every polynomial can be factorized into linear factors.

For convenience, we use the notation

\[ A X := \{ Ax \mid x \in X \}. \]

A subspace \( X \) is called invariant under \( A \) if \( AX \subseteq X \). We also call a linear map a linear operator, since the underlying physical vector spaces could have a much broader meaning than merely vector spaces.

The basic question in linear algebra here is to find a basis such that the matrix of a given map under the basis is as diagonal as possible. That is to say, we want to find a finest decomposition

\[ V = \bigoplus X_i, \quad AX_i \subset X_i. \]

Here finest means that \( X_i \) has as small dimension as possible. Since \( AX_i \subset X_i \), the study of \( A \) on \( V \) is then reduced to the study of \( A \) on each \( X_i \), which should be significantly smaller than \( V \).

Clearly, the best scenario would be \( \dim(X_i) = 1 \). In this case, \( X_i = \text{span}\{x\} \) for some \( x \) and since \( Ax \in X_i \), there exists a scalar \( \lambda \) such that

\[ Ax = \lambda x. \]

If \( \lambda \in \mathbb{F} \) and \( x \in V \setminus \{0\} \) satisfies \( Ax = \lambda x \), we call \( \lambda \) an \textbf{eigenvalue}, \( x \) an \textbf{eigenvector} associated with \( \lambda \), and \( (\lambda, x) \) an \textbf{eigenpair}. A problem of finding eigenpairs is called an eigenvalue problem.

If we can find eigenpairs \( (\lambda_1, u_1), \ldots, (\lambda_n, u_n) \) of a map \( A \) such that \( \{u_1, \ldots, u_n\} \) is a basis, then the linear map \( A \) can be describes by \( A(x^t u_i) = \sum \lambda_i x^t u_i \), and the corresponding matrix of the map \( A \) under the basis is \( \text{diag}(\lambda_1, \ldots, \lambda_n) \).

There are cases where we cannot find such a basis; namely, there do not exist \( n \) linearly independent eigenvectors. The spectral theory tells us to look for \textbf{generalized} eigenvectors, namely, solutions to

\[ (A - \lambda)^k x = 0. \]

The spectral theory says that we can find a basis that consists of eigenvectors and generalized eigenvectors.

As we have seen before, a good representation of a linear operator allows us to find quickly many other related linear operators such as \( e^A \), \( \sin A \), etc.
4.1 Examples of Linear Operators

Example 4.1. Let $\mathbb{F} = \mathbb{R}$ and $V_n$ be the set of all polynomials of degree $\leq n$. If we denote by $e_i$ the $i$th power function $t \in \mathbb{R} \rightarrow t^i$, then $V_n = \text{span}\{e_0, \ldots, e_n\}$. Set $V = \bigcup_{n=1}^{\infty} V_n$. Let $A : V \rightarrow V$ be the map defined by

$$Af = f(0)e_0 + bf(1)e_1$$

for all $f \in V$ where $b \in \mathbb{R}$ is a fixed number and $f(t)$ is the value of $f$ at $t$. It is easy to show that $A$ is a linear map from $V_n$ to $V_n$ itself, as long as $n \geq 1$.

(1) Let $u_i = e_i - e_1$ for $i \geq 2$. Then $u_i(0) = u_i(1) = 0$ so that

$$Au_i = 0.$$ 

For $n \geq 2$, set $U = \text{span}\{u_2, \ldots, u_n\}$. Then

$$V_n = V_1 \oplus U, \quad AV_1 \subset V_1, \quad AU = \{0\}.$$ 

Hence, to study $A$, we need only concentrate our attention to the behavior of $A$ on $V_1$.

(2) Now consider $A$ on $V_1$. Under the basis $\{e_0, e_1\}$ its associated matrix is

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

since $A(e_0, e_1) := (Ae_0, Ae_1) = (e_0, e_1)A$.

Suppose $b \neq 1$. Then we can find two eigenpairs, $(1, (b - 1)e_0 - be_1)$ and $(b, e_1)$. Hence, under the basis $\{(b - 1)e_0 - be_1, e_1\}$ the matrix associated with $A$ takes the form

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

Suppose $b = 1$. Then we can find only one eigenpair $(1, e_1)$, all the others are of the form $(1, ce_1)$ ($c \neq 0$), so that we cannot find a basis consisting of eigenvectors.

On the other hand, denoting $I$ the identity map, we have

$$(A - I)e_0 = e_1, \quad (A - I)e_1 = 0, \quad (A - I)^2 f = 0 \quad \forall f \in V_1.$$ 

Example 4.2. With same vector space as in the previous example, consider the linear differential operator

$$B = \frac{d}{dt} + I : f \in V \rightarrow f' + f.$$ 

It is easy see that $B$ is a linear map from $V_n$ to $V_n$ itself, for each $n \geq 0$.

If we focus our attention to $V_1$. The corresponding matrix under the basis $\{e_1, e_0\}$ is

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

This is the same matrix as in the previous example with $b = 1$. It is important to note that in the previous case the basis is $\{e_0, e_1\}$ whereas in the current case the basis is $\{e_1, e_0\}$. 

124

CHAPTER 4. SPECTRAL THEORY
4.1. EXAMPLES OF LINEAR OPERATORS

In $V_n$ ($n \geq 2$), the matrix of the linear operator $B$ under the basis $\{e_0, \cdots, e_n\}$ is

$$
\begin{pmatrix}
1 & 1 & 0 & \cdots & 0 \\
0 & 1 & 2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \cdots & 0 \\
0 & 0 & \cdots & 1 & n \\
0 & 0 & \cdots & 0 & 1
\end{pmatrix}
$$

If we use the basis $\{\frac{1}{\pi}e_n, \cdots, \frac{1}{\pi}e_0\}$, the matrix becomes

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

Example 4.3. Consider $\mathbb{F} = \mathbb{R}$, $W = \text{span}\{\sin \pi t, \cos \pi t\}$, and the differential operator $B := \frac{d}{dt} + I$. It is a linear operator from $W$ to itself. Under the basis $\{\sin \pi t, \cos \pi t\}$, the matrix of $B$ is

$$
\begin{pmatrix}
1 & -\pi \\
\pi & 1
\end{pmatrix}
$$

One can show that this matrix does not have any real eigenvalues. Indeed, the eigenvalues are $1 \pm \pi i$.

If $\mathbb{F} = \mathbb{C}$, then under the basis $\{\sin(\pi t) + i \cos(\pi t), \sin(\pi t) - i \cos(\pi t)\}$ the matrix becomes

$$
\begin{pmatrix}
1 + \pi i & 0 \\
0 & 1 - \pi i
\end{pmatrix}
$$

Example 4.4. Let $U = M_2(\mathbb{R})$, a four dimensional vector space. Fix a matrix $A = (a_{ij}) \in U$, we can define a linear operator $A : B \in U \to AB \in U$. Let

$$
e_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad e_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.
$$

Then $\{e_1, \cdots, e_4\}$ is a basis of $U$. Under this basis, the matrix associated with the linear operator $A$ is

$$
\begin{pmatrix}
a_{11} & 0 & a_{12} & 0 \\
0 & a_{11} & 0 & a_{12} \\
a_{21} & 0 & a_{22} & 0 \\
0 & a_{21} & 0 & a_{22}
\end{pmatrix}
$$

If we use the basis $\{e_1, e_3, e_2, e_4\}$ the matrix of the operator $A$ becomes

$$
\begin{pmatrix}
a_{11} & a_{12} & 0 & 0 \\
a_{21} & a_{22} & 0 & 0 \\
0 & 0 & a_{11} & a_{12} \\
0 & 0 & a_{21} & a_{22}
\end{pmatrix}
$$
Example 4.5. Let $V = C^2_0([0, \pi])$ be the space of all twice continuous differentiable functions on $[0, 2\pi]$ that vanish at 0 and $\pi$. Consider the linear operator $A = \frac{d^2}{dx^2}$. The eigenvalue problem is to find $\lambda \in \mathbb{C}$ and $u \in V$ such that

$$u''(x) = \lambda u(x) \quad \forall x \in (0, \pi), \quad u(0) = u(\pi) = 0.$$ 

One can obtain a complete set of eigenpairs $\{-n^2, \sin(nx)\}_{n=1}^{\infty}$.

If one uses $\{\sin(nx)\}_{n=0}^{\infty}$ as the basis for $V$, then the operator $\frac{d^2}{dx^2}$ can be written as

$$A \sum_{n=0}^{\infty} a_n \sin(nx) = \frac{d^2}{dx^2} \sum_{n=0}^{\infty} a_n \sin(nx) = -\sum_{n=0}^{\infty} n^2 a_n \sin(nx).$$

Also, for each integer $n \geq 1$ and real number $t$,

$$e^{tA} \sin(nx) = \sum_{k=0}^{\infty} \frac{1}{k!} (tA)^k \sin(nx) = \sum_{k=1}^{\infty} \frac{(-n^2 t)^k}{k!} \sin(nx) = e^{-n^2 t} \sin(nx).$$

Now we consider the differential equation $w_t = Aw$ for $w : t \in [0, \infty) \rightarrow V(t) \in V$. Supplied with an initial data $w(0) = u_0 = \sum a_n \sin(nx) \in V$, the solution is

$$w(t) = e^{tA} u_0 = \sum e^{-n^2 t} a_n \sin(nx).$$

Here, one first compute the solution in a formal level, then uses a rigorous mathematics to verify the result.

Exercise 1. For the operator in example 4.1, find the matrix associated with the operator $A$ under the basis $\{e_0, e_1, \cdots, e_n\}$ for $V_n$.

Exercise 2. Assume that $V = \bigoplus_{i=1}^{m} X_i$ and $AX_i \subset X_i$. Let $U_i = \{x_i^1, \cdots, x_i^{n_i}\}$ be a basis for $X_i$, $i = 1, \cdots, m$. Show that under the basis $\{U_1, \cdots, U_m\}$, the matrix associated with the map $A$ takes the form

$$
\begin{pmatrix}
A_1 & 0 & \cdots \\
0 & A_2 & \cdots \\
\vdots & \ddots & \ddots
\end{pmatrix}
$$

where $A_i$ is an $n_i \times n_i$ matrix associated with the map $A$ on $X_i$.

Exercise 3. Suppose $A$ is the matrix of $A$ under the basis $\{u_1, \cdots, u_n\}$. What is the matrix $B$ of $A$ under the basis $\{u_n, \cdots, u_1\}$? Does $B$ have any relation with $A^T$? If we write $B = P^{-1}AP$, what is $P$?

Exercise 4. In example 4.4, find the matrix for the linear map: $B \in U \rightarrow BA \in U$. Extend the result to the case $U = M_2(\mathbb{F})$.

Exercise 5. Carry out the same formal analysis as in example 4.5 for $A = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ on the space $V$ of smooth functions on $[0, \pi] \times [0, \pi]$ with zero boundary data. Also, find solutions to initial value problem $w_t = Aw$ for $w : t \in [0, \infty) \rightarrow V$ with initial data $w(0) = w_0 = \sum_{m,n} a_{mn} \sin(nx) \sin(my)$.
4.2 Functions of Matrices

When $A$ is a square matrix, we can form a polynomial of the matrix

$$p(A) = a_0I + a_1A + \cdots + a_nA^n.$$ 

For real or complex matrices, we can also define transcendental functions:

$$\exp(A) = \sum_{n=1}^{\infty} \frac{1}{n!} A^n,$$

$$\sin(A) = \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k+1)!} A^{2k+1},$$

$$\exp(iA) = \cos A + i \sin A = \sum_{n=1}^{\infty} \frac{1}{n!} (iA)^n,$$

$$\ln(I + A) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k + 1} A^{2k+1}.$$

For diagonal matrices, the calculation is very simple. Suppose

$$\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$$

Then

$$\Lambda^k = \text{diag}(\lambda_1^k, \cdots, \lambda_n^k),$$

$$\exp(\Lambda) = \text{diag}(e^{\lambda_1}, \cdots, e^{\lambda_n}).$$

Approximate any continuous function $f : \mathbb{R} \to \mathbb{R}$ by polynomials, we also can extend it to be functions of square matrices:

$$f(\Lambda) = \text{diag}(f(\lambda_1), \cdots, f(\lambda_n)).$$

For general square matrix $A$, the direct computation of these functions from their series expansion is time consuming. However, if $A$ is diagonalizable, the computation will be very easy if we diagonalize it. Suppose

$$A = P^{-1}\Lambda P$$

Then, since $A^k = PA^kP$,

$$\exp(A) = \sum_{i=0}^{\infty} \frac{1}{i!} A^n = P^{-1}e^\Lambda P,$$

$$f(A) = P^{-1}f(\Lambda)P.$$ 

Exercise 6. (i) Using the series expansion show that for any square matrix $A$,

$$\frac{d}{dt}\exp(tA) := \lim_{\tau \to 0} \frac{e^{(t+\tau)A} - e^{tA}}{\tau} = Ae^{tA} = e^{tA}A.$$ 

(ii) Show that for any $A \in M_n(\mathbb{C})$, $B \in M_{m,n}(\mathbb{C})$ and $C \in M_{p,n}(\mathbb{C})$, the functions $X(t) = e^{tA}B$ and $Y(t) = Ce^{tA}$ solve the system of differential equations

$$\frac{dX}{dt} = AX, \quad \frac{dY}{dt} = YA.$$ 

(iii) For $A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$ work out the details of (i) and (ii). In particular, (a) write down the system of equations, in scalar form for $X = \begin{pmatrix} x \\ y \end{pmatrix}$ and similarly for $Y = \begin{pmatrix} u \\ v \end{pmatrix}$. (b) write down, in scalar form, the general solutions you get.
4.3 Spectrum

The eigenvalue problem can be written as

\[(A - \lambda I)x = 0 \quad \text{or} \quad (A - \lambda I)x = 0, \tag{4.1}\]

where \(A\) is the matrix of the operator \(A\) under a basis, say \(\{e_1, \ldots, e_n\}\), and \(x \in \mathbb{F}^n\) is the coordinates of \(x\) under the basis.

Using determinants, we know that a necessary and sufficient condition for \(\lambda\) to be an eigenvalue is

\[\det(A - \lambda I) = 0.\]

**Definition 4.1.** (1) Let \(A \in M_n(\mathbb{F})\). The **characteristic polynomial** of \(A\) is

\[p_A(z) := \det(zI - A) = z^n - \text{Tr}(A)z^{n-1} + \cdots + (-1)^n\det(A).\]

(2) The **spectrum** of \(A\) is the set

\[\sigma(A) := \{\lambda \in \mathbb{F} \mid p_A(\lambda) = 0\}\]

An element \(\lambda \in \sigma(A)\) is called an **eigenvalue** of \(A\), and its **multiplicity** as eigenvalue is its multiplicity as the root to the equation \(p_A(\lambda) = 0\).

In the sequel, a polynomial is called **monic** if its highest order term has coefficient 1. A monic polynomial cannot be identically zero.

To define the characteristic polynomial of a linear operator, we need

**Lemma 4.1.** If \(B\) is similar to \(A\), i.e., \(B = P^{-1}AP\) for some invertible matrix \(P\), then their characteristic polynomials are identical: \(p_A(\lambda) = p_B(\lambda)\).

**Definition 4.2.** (1) Let \(A\) be a linear operator mapping an \(n\)-dimensional vector space \(V\) into itself. Let \(A\) be the corresponding matrix of \(A\) under an arbitrarily chosen basis. Then \(p_A(\lambda)\) is called the **characteristic polynomial** of the map \(A\), written as

\[p_A(z) = z^n + a_1z^{n-1} + \cdots + a_n.\]

Here \(-a_1\) is called the **trace** of \(A\) and \((-1)^n a_n\) the determinant of \(A\).

(2) Similarly, \(\sigma(A)\) is called the **spectrum** of \(A\), written as \(\sigma(A)\).

(3) For \(\lambda \in \sigma(A)\), the set

\[E(\lambda) := \bigcup_{i=1}^{\infty} \ker\left((\lambda I - A)^i\right)\]

is called the **eigenspace** of \(A\) associated with the eigenvalue \(\lambda\). Here \(\ker(B)\), the **kernel** of \(B\), is the set

\[\ker(B) := \{x \in V \mid Bx = 0\}.\]

One notices from the Lemma that \(p_A(\lambda)\) is well-defined. It is intrinsic, in the sense that \(p_A(\lambda)\) is independent of the choices of bases for the underlying vector space. We leave it as an exercise to show that an eigenspace is a subspace.

In the infinite space dimensional cases, characteristic polynomial of an operator is not defined, neither are the trace and determinant. Nevertheless, the spectrum and eigenspace of an operator mapping a vector space into itself is well-defined.
4.3. SPECTRUM

Corollary 4.1. Given a linear operator \( A \) from a finite dimensional vector space to itself,

\[
\sigma(A) = \mathbb{F} \setminus \rho(A), \quad \rho(A) := \{ \lambda \in \mathbb{F} \mid \lambda I - A \text{ is an isomorphism} \}.
\]

In addition, \( E(\lambda) = \{0\} \) if \( \lambda \in \rho(A) \).

Remark 4.1. One notices that the above formula does not involve any coordinate systems. Indeed, it can be used as the definition of spectrum for linear operators that map an infinite dimensional vector space into itself. For normed vector spaces, one needs a stronger condition for \( \rho(A) \):

\[
\rho(A) := \{ \lambda \in \mathbb{C} \mid \lambda I - A \text{ is an isomorphism and } (\lambda I - A)^{-1} \text{ is bounded} \}.
\]

Then the set \( \sigma(A) := \mathbb{C} \setminus \rho(A) \) consists of two parts: The part \( \sigma_p(A) \), called point spectrum, in which each \( \lambda \) admits at least one eigenvector, and the part \( \sigma_c(A) \), called continuum spectrum in which \( \lambda \) does not admits any eigenvector, namely \( (\lambda I - A) \) is one-to-one, but \( (\lambda I - A)^{-1} \) is not bounded.

The following result will be proven later.

Lemma 4.2. If \( p(z) = \prod_{i=0}^{n}(z - \lambda_i) \) is the characteristic polynomial of a matrix \( A \in M_n(\mathbb{F}) \), then for any polynomial \( f(s) = \sum_{i=0}^{m} a_i s^i \) with coefficients in \( \mathbb{F} \), the characteristic polynomial of \( f(A) := \sum_{i=1}^{m} a_i A^i \) is \( p_f(z) = \prod_{i=1}^{n}(z - f(\lambda_i)) \).

Example 4.6. Suppose the characteristic polynomial for \( A \in M_3(\mathbb{R}) \) is \( p(z) = z(z^2 + 1) \). What is the characteristic polynomial for \( A^2, e^A, f(A) \) where \( f : \mathbb{C} \to \mathbb{C} \) is analytic (i.e., \( f(z) = \sum c_n z^n \) is uniformly convergent in \( z \in \mathbb{C} \)).

Solutions. The eigenvalues of \( A \) are 0, i, and -i. Hence, the eigenvalues of \( f(A) \) is \( f(0), f(i), f(-i) \). Thus, the characteristic polynomial for \( f(A) \) is \( p_f(z) = (z - f(0))(z - f(i))(z - f(-i)) \). In particular,

\[
p_{A^2}(z) = z(z + 1)^2, \quad p_{A^3}(z) = z(z^2 + 1), \quad p_{e^A}(z) = (z - 1)(z - e^1)(z - e^{-1}) = z(z^2 - 2z \cos 1 + 1).
\]


Exercise 8. Show that an eigenspace of a linear map is a subspace. Also show that in an \( n \)-dimensional vector space, there exists \( k \leq n \) such that

\[
E(\lambda) = \ker \left((\lambda I - A)^k\right).
\]

Exercise 9. Suppose the characteristic polynomial of \( A \in M_3(\mathbb{R}) \) is \( p(z) = z^3 - 2z^2 + z \). What is the characteristic polynomial for \( A^2, A^3, e^A, \sin A, f(A) \)?

Exercise 10. Fix a matrix \( A \in M_3(\mathbb{F}) \). We define a linear operator \( \Lambda : M_3(\mathbb{F}) \to M_3(\mathbb{F}) \) by \( \Lambda(B) = AB \) for all \( B \in M_3(\mathbb{F}) \). Find the characteristic polynomial of the map \( \Lambda \).

Exercise 11. Let \( F = \mathbb{R} \) and \( V_n \) be the set of polynomials of degree \( \leq n \). Let \( \Phi = \frac{d^2}{dx^2} + 2\frac{d}{dx} + I \) be the operator mapping \( f \in V_n \) to \( \Phi(f) = f'' + 2f' + f \). Find the characteristic polynomial of \( \Phi \) for \( V_1, V_2, \) and \( V_3 \) respectively.
4.4 The Cayley-Hamilton Theorem

Given a matrix $A \in M_n(\mathbb{F})$, if we regard $M_n(\mathbb{F})$ as a $n^2$ dimensional vector space over $\mathbb{F}$, then the $n^2 + 1$ matrices

$$I, A, A^2, \ldots, A^{n^2}$$

are linearly dependent. Hence, there is a non-trivial linear combination that is equal to $0I$. The following theorem says that indeed such a combination can be obtained by using only the first $n + 1$ matrices in the list.

**Theorem 4.1. (Cayley-Hamilton Theorem)**

Let $A$ be a linear operator from an $n$-dimensional vector space to itself and $p(z) = z^n + a_1 z^{n-1} + \cdots + a_n$ be the characteristic polynomial of $A$. Then

$$p(A) := A^n + a_1 A^{n-1} + \cdots + a_n I = 0I.$$

We remark that $\det(zI - A)$ is a scalar, whereas $p(A)$ is a matrix, so that $p(B) \neq \det(BI - A)$. For example, let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then its characteristic polynomial is

$$p(z) = \det(zI - A) = z^2 - (a + d)z + (ad - bc).$$

The Cayley-Hamilton theorem says that there is the identity $p(A) = 0I$, i.e.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} - (a + d) \begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad - bc) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

**Proof.** Fix a basis and let $A$ be the matrix of $A$ under the basis. Let $Q(z) = (q_{ij}(z))_{n \times n}$ be the cofactor matrix of $zI - A$, i.e.

$$q_{ij}(z) = (-1)^{i+j} \det((zI - A)_{ji})$$

where $(zI - A)_{ji}$ is the $n-1$ by $n-1$ matrix obtained by deleting from $zI - A$ the jth row and ith column. By using the Laplace expansion, we have

$$Q(z)(zI - A) = \det(zI - A)I = p(z)I.$$

Writing $Q(z) = Q_0 + zQ_1 + \cdots + z^{n-1}Q_{n-1}$ we then have

$$(Q_0 + zQ_1 + \cdots + z^{n-1}Q_{n-1})(zI - A) = (z^n + a_1 z^{n-1} + \cdots + a_n)I$$

Since $\mathbb{F}$ has infinitely many elements, the coefficients on both sides have to be the same. Hence,

$$Q_{n-1} = I, \quad Q_{n-2} - Q_{n-1} A = a_1 I, \quad \ldots, \quad -Q_0 A = a_n I.$$

This implies that

$$p(A) = A^n + a_1 IA^{n-1} + \cdots + a_n I$$

$$= Q_{n-1} A^n + (Q_{n-2} - Q_{n-1} A)A^{n-1} + \cdots + (-Q_0 A) = 0I. \quad \Box$$
4.4. THE CAYLEY-HAMILTON THEOREM

Definition 4.3. Let $A$ be a linear operator from a vector space $V$ to itself. A monic polynomial $q$ of minimum degree that annihilates $A$ (e.g. $q(A) = 0I$) is called the minimal polynomial of $A$.

We point out that the minimal polynomial is unique. Indeed, if $q_1$ and $q_2$ are polynomials satisfying $q_1(A) = 0I = q_2(A)$, then by the following lemma their greatest common divisor $r(z)$ also satisfies $r(A) = 0I$. In particular, the minimal polynomial is a factor of the characteristic polynomial.

Lemma 4.3. (1) Suppose $q_1(z), q_2(z)$ are polynomials with coefficients in a field $\mathbb{F}$, not necessarily closed. If $r(z) = (q_1(z), q_2(z))$ is the greatest common divisor, then there exist polynomials $a(z)$ and $b(z)$ such that

$$a(z)p(z) + b(z)q(z) = r(z).$$

Proof. We use a degree reduction process. If $\deg(p) = \deg(r)$, then set $b = 0$ and $a = 1/p_0$ where $p_0$ is the leading coefficient of $p$, and we are done. If $\deg(p) > \deg(r)$, we find the remainder $\tilde{q}$ of $q \div p$ so that $\tilde{q} = q - ap$, $\deg(\tilde{q}) < \deg(p)$, and the greatest common divisor $(\tilde{q}, p)$ is still $r$. If $\deg(\tilde{q}) > \deg(r)$, we replace $(p, q)$ by $(\tilde{q}, p)$ and continue the previous step; otherwise we must have $\tilde{q} = q_0r$ where $q_0$ is the leading coefficient of $\tilde{q}$. After escrowing the remainder equations, we obtain $r = ap + bq$. \hfill \Box

Example 4.7. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{F})$. Find the minimal polynomial of $A$.

Solution. Set $\alpha = \text{Tr}(A) = a + d$ and $\beta = \text{det}(A) = ad - bc$. Then the characteristic polynomial of $A$ is

$$p(z) = z^2 - \alpha z + \beta.$$ 

The minimal polynomial is $p(z)$, except the case $A = aI$ in which the minimal polynomial is $q(z) = z - a$.

Example 4.8. Find $A^5 + 2A^4 + A^3 + A^2 + A + I$ where $A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$.

Solution. The characteristic polynomial of $A$ is $p(z) = z^2 + 1$. Noting that $q(z) := z^5 + 2z^4 + z^3 + z^2 + z + 1 = (z^2 + 1)(z^3 + 2z^2 - 1) + z + 2$. It follows from Cayley-Hamilton theorem that $q(A) = A + 2I = \begin{pmatrix} 3 & -2 \\ 1 & 1 \end{pmatrix}$.

Exercise 12. (1) Let $P_1, \cdots, P_m, Q_1, \cdots, Q_p$ be $n \times n$ matrices, and

$$P(z) = \sum_j P_j z^j, \quad Q(z) = \sum_k Q_k z^k, \quad R(z) = \sum_i \left( \sum_{j+k=i} P_j Q_k \right) z^i.$$ 

Suppose an $n \times n$ matrix $A$ commutes with each of $Q_1, \cdots, Q_p$, show that $R(A) = P(A)Q(A)$.

(2) For $n = 2$, find an example of $P$ and $Q$ such that $R(B) \neq P(B)Q(B)$ for some $n \times n$ matrix $B$.

Exercise 13. Show that in the field $\mathbb{C}$, the minimal polynomial of a real square matrix has real coefficients.

Exercise 14. Show that in the finite dimensional case, a minimal polynomial of a linear operator exists and is unique.

Exercise 15. Find the minimal polynomials of the following matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
Exercise 16. For $A = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}$ find $A^5 + 2A^4 + 3A^3 + 4A^2 + 5A + 6I$.

Exercise 17. Show that if two matrices are similar, then their minimal polynomials are identical.

Exercise 18. Suppose the characteristic polynomial of $A$ is $p(z) = \prod_i (z - \lambda_i)^{n_i}$ where $\lambda_1, \cdots, \lambda_m$ are distinct. Show that the minimal polynomial $q$ of $A$ has the form $q(z) = \prod_i (z - \lambda)^{m_i}$ with $1 \leq m_i \leq n_i$.

Exercise 19. Given a monic polynomial $p(z) = z^n - a_1t^{n-1} + \cdots - a_n$, show that the minimal and characteristic polynomials of following matrix are the same and equal to $p$:

$$
\begin{pmatrix}
0 & 0 & \cdots & 0 & a_{n-1} \\
1 & 0 & \cdots & 0 & a_{n-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \ddots & 0 & a_1 \\
0 & 1 & \cdots & 1 & a_0
\end{pmatrix}
$$

Exercise 20. Suppose the minimal polynomial of $A \in M_5(\mathbb{R})$ is $p(z) = (z - 1)^2(z^2 + 1)$. What is the characteristic polynomial for $A$, $A^2$, $A^3$, $A^5$, $e^A$, $\sin A$?
4.5 The Spectral Theorem

Theorem 4.2. (The Spectral Theorem)

Let $A$ be a linear operator from a finite dimensional vector space $V$ over a closed field to $V$ itself. Then

$$V = \bigoplus_{\lambda \in \sigma(A)} E(\lambda), \quad E(\lambda) := \bigcup_{k \geq 1} \ker((\lambda I - A)^k), \quad AE(\lambda) \subset E(\lambda)$$

That is, $V$ is a direct sum of eigenspaces which are invariant under $A$; in other words, every vector can be written as a sum of eigenvectors, genuine or generalized.

We remark that the theorem may not be true when the field is not closed. That's why in some cases, we need to extend vector spaces over the field $\mathbb{R}$ to the field $\mathbb{C}$.

The spectral theorem can be extended to infinite dimensional normed vector spaces over $\mathbb{C}$ in which the finiteness condition on space dimension is replaced by a requirement that the operator $A$ be a compact operator, i.e., the set $\{Ax \mid \|x\| = 1\}$ is compact.

We shall use the following lemma to prove the theorem.

Lemma 4.4. Suppose $p_1(z), \ldots, p_k(z)$ are polynomials that are pairwise prime to each other, i.e. the greatest common divisor $(p_i, p_j) = 1$ for all $i \neq j$. Then

$$\ker \left( \prod_i p_i(A) \right) = \bigoplus_i \ker(p_i(A)).$$

Proof. First we consider the case $k = 2$. Let $a(z)$ and $b(z)$ be polynomials such that $1 = a(z)p_1(z) + b(z)p_2(z)$. Then $I = a(A)p_1(A) + b(A)p_2(A)$ so that for each $x \in V$, we can write

$$x = x_2 + x_1, \quad x_2 := a(A)p_1(A)x, \quad x_1 := b(A)p_2(A)x.$$

Now if $x \in \ker(p_1(A)p_2(A))$, then $p_1(A)x_1 = b(A)p_2(A)p_1(A)x = 0$ so that $x_1 \in \ker(p_1(A))$. Similarly, $x_2 \in \ker(p_2(A))$. As $\ker(p_1(A)) \subset \ker(p_1(A)p_2(A))$ for $i = 1, 2$, we see that

$$\ker(p_1(A)p_2(A)) = \ker(p_1(A)) + \ker(p_1(A)).$$

To show that the sum is direct, let $x \in \ker(p_1(A)) \cap \ker(p_2(A))$, then $p_1(A)x = 0 = p_2(A)x$ so that $x = a(A)p_1(A)x + b(A)p_2(A)x = 0$. Hence, $\ker(p_1(A)) \cap \ker(p_2(A)) = \{0\}$. Thus, $\ker(p_1(A)p_2(A)) = \ker(p_1(A)) \oplus \ker(p_1(A))$.

Now in the general case $k \geq 3$, we can use the fact that $p_j$ is prime to $p_{j+1} \cdots p_k$ for $j = 1, \ldots, k-1$ to obtain

$$\ker(p_1(A) \cdots p_k(A)) = \ker(p_1(A)) \oplus \ker(p_2(A) \cdots p_k(A)) = \ker(p_1(A)) \oplus \left( \ker(p_2(A)) \oplus \ker(p_3(A) \cdots p_k(A)) \right) = \cdots = \ker(p_1(A)) \oplus \ker(p_2(A)) \oplus \cdots \oplus \ker(p_k(A)).$$

Proof of the Spectral Theorem. Let $p(z)$ be the characteristic polynomial of $A$. By the Cayley-Hamilton theorem, $p(A) = 0I$, i.e. $p(A)x = 0$ for all $x \in V$, so that $V = \ker(p(A))$. As $F$ is closed, we can write $p(z) = \prod_{i=1}^k (z - \lambda_i)^{n_i}$ where $\lambda_i \neq \lambda_j$ for $i \neq j$. The assertion of the theorem then follows from the lemma by setting $p_i(z) = (z - \lambda_i)^{n_i}, i = 1, \ldots, k$. $\Box$
Exercise 21. (1) Let $\mathbb{Z}^2 = \{m + ni \mid m, n \in \mathbb{Z}\}$, with summation and multiplication as that of complex numbers. Show that in $\mathbb{Z}^2$, $1 + 2i$ is prime to both $2 + i$ and $2 - i$. However, $1 + 2i$ is not prime to $(2 + i)(2 - i) = 5$ since $5 = (1 + 2i)(1 - 2i)$.

(2) Let $p_1(z), \ldots, p_n(z)$ be polynomials with coefficients in a field. Show that if $p_1(z)$ is prime to each of $p_2(z), \ldots, p_k(z)$, then $p_1(z)$ is prime to $p_2(z) \cdots p_k(z)$.

Exercise 22. Let $p(z) = z, q(z) = (z - 1)^2$, and $r(z) = (z - 2)^3$. Find polynomials $a(z), b(z)$, and $c(z)$ such that $a(z)p(z) + b(z)q(z) + c(z)r(z) = 1$.

Exercise 23. Let $\{e_1, e_2, e_3\}$ be the basis of a vector space and $A$ be a linear operator defined by

$$A(\sum_i x^i e_i) = (x^1 + 2x^2 + 3x^3)e_1 + (2x^3 + x^3)e_2 + 5x^3 e_3 \quad \forall (x^1, x^2, x^3) \in \mathbb{F}_n.$$ 

Find the spectral decomposition.
4.6 Nilpotent Operator

Now we study the action of $A$ on its eigenspace $E(\lambda)$. By working on $A - \lambda I$, we can assume that $\lambda = 0$. Since $E(0)$ is invariant, we can assume that $E(0) = V$.

**Definition 4.4.** A linear operator from a vector space $V$ to itself is called **nilpotent** if $A^n = 0I$ for some positive integer $n$.

Let $A$ be a nilpotent operator over a vector space $V$. Let $k$ be the positive integer such that

$$A^{k+1} = 0I \neq A^k.$$

Then

$$V = \ker(A^{k+1}) \supseteq \ker(A^k) \supseteq \cdots \supseteq \ker(A^0) = \{0\}.$$ 

Here the first inequality follows from $A^k \neq 0I$, whereas the rest inequalities follow from the following:

**Lemma 4.5.** For each $j = 1, \cdots, k$, $A$ maps $\ker(A^{j+1})/\ker(A^j)$ injectively to $\ker(A^j)/\ker(A^{j-1})$. Consequently, if define

$$n_j := \dim\left(\ker(A^{j+1})/\ker(A^j)\right), \quad j = 0, 1, \cdots, k.$$ 

then $n_k \leq n_{k-1} \leq \cdots \leq n_0$. In addition,

$$\dim(\ker(A^{j+1})) = n_j + \dim(A^j), \quad \forall j$$

$$\dim(V) = \dim(\ker(A^{k+1})) = n_k + n_{k-1} + \cdots + n_0.$$ 

Proof. If $x \in \ker(A^{j+1})$, $A^{j}(Ax) = 0$, i.e., $Ax \in \ker(A^{j})$. Hence, $A$ maps $\ker(A^{j+1})$ to $\ker(A^{j})$.

Next, suppose $y \equiv x \mod(\ker(A^{j}))$. Then writing $y = x + z$ where $z \in \ker(A^{j})$, we have $Ay = Az \in \ker(A^{j-1})$, i.e., $Ay \equiv Ax \mod(\ker(A^{j-1}))$. Thus, $A$ maps $\ker(A^{j+1})/\ker(A^j)$ to $\ker(A^j)/\ker(A^{j-1})$.

Finally, we show that the map is injective. Suppose $Ax \equiv 0 \mod(\ker(A^{j-1}))$. Then $0 = A^{j-1}(Ax) = A^jx$ so that $x \equiv 0 \mod(\ker(A^j))$. Thus, the map is injective. 

Now we construct a basis for $V$ as follows.

Select $x_1^k, \cdots, x_n^k$ such that $x_i^k \mod(\ker(A^k))$, $i = 1, \cdots, n_k$, form a basis for $\ker(A^{k+1})/\ker(A^k)$. Then

$$V = \ker(A^{k+1}) = \left( \oplus_{i=1}^{n_k} \text{span}\{x_i^k\} \right) \oplus \ker(A^k).$$ 

Suppose we have constructed a basis $x_i^j \mod(\ker(A^j))$, $i = 1, \cdots, n_j$ for $\ker(A^{j+1})/\ker(A^j)$. Set

$$x_i^{j} = Ax_i^{j-1}, \quad i = 1, \cdots, n_j.$$ 

Then by the lemma, these vectors are linearly independent in $\ker(A^j)/\ker(A^{j-1})$. Hence, we can expand it to $x_i^{j-1} \mod(A^{j-1})$, $i = 1, \cdots, n_{j-1}$ to become a basis for $\ker(A^j)/\ker(A^{j-1})$. There holds

$$\ker(A^j) = \left( \oplus_{i=1}^{n_{j-1}} \text{span}\{x_i^{j-1}\} \right) \oplus \ker(A^{j-1}).$$

Recursively finding these bases for $j = k, k-1, \cdots, 1$, we then obtain $\{x_i^j \mid i = 1, \cdots, n_j, j = k, k-1, \cdots, 0\}$ which is a basis for $V$ since $\ker(A^0) = \{0\}$. 

Since $x_i^{j-1} = Ax_i^j$ for $i \leq n_j$, we can list all the elements in the basis of $V$ as follows:

$$
\begin{align*}
&x_1^k, \ldots, x_{n_k}^k, \\
&Ax_1^k, \ldots, Ax_{n_k}^k, x_{n_k+1}^{k-1}, \ldots, x_{n_{k-1}}^{k-1}, \\
&\vdots \quad \vdots \quad \vdots \\
&A^kx_1^k, \ldots, A^kx_{n_k}^k, A^{k-1}x_{n_k+1}^{k-1}, \ldots, A^{k-1}x_{n_{k-1}+1}^{k-1}, \ldots, x_0^{n_0+1}, \ldots, x_0^{n_0}.
\end{align*}
$$

We can order them in the vertical direction. For each $i = 1, \ldots, n_0$, there is an integer $k_i$ such that $x_i^j = A^{k_i-j}x_i^{k_i}$ for $j = 0, \ldots, k_i - 1$ and $A^{k_i+1}x_i^k = 0$. Denoting $n_0 = m$, and $x_i^{k_i} = x_i$, we then have the following.

**Theorem 4.3.** Let $A$ be a nilpotent operator mapping a finite dimensional vector space into itself. Then there exists a basis of the form

$$
\{A^{k_1}x_1, A^{k_2}x_1, \ldots, x_1, A^{k_2}x_2, A^{k_2}x_2, \ldots, x_2, \ldots, A^{k_m}x_m, \ldots, x_m\}
$$

where $A^{k_i+1}x_i = 0$ for $i = 1, \ldots, m$. In particular, under this basis, the matrix $A$ associated with the operator is given by the Jordan form

$$
A = \begin{pmatrix} 
J_{k_1} & 0 \\
\vdots & \ddots \\
0 & \ddots & J_{k_m} 
\end{pmatrix}, \quad \text{where } J_k := \begin{pmatrix} 
0 & 1 & 0 \\
0 & 1 & \ddots \\
0 & 0 & 1 
\end{pmatrix}_{k \times k}
$$

where $k = k_1 \geq k_2 \geq \cdots \geq k_m \geq 1$.

Exercise 24. Show that in finite dimensional case, a linear operator $A$ is nilpotent if and only if $\sigma(A) = \{0\}$.

Exercise 25. When the space dimension is 5, list all the possible Jordan matrices for nilpotent operators. Group them according to the similarity of matrices.

Exercise 26. Suppose $A \in M_6(\mathbb{R})$ satisfies $A^2 = 0I$. List all possible Jordan normal forms for $A$.

Exercise 27. Let $V$ be the space of all polynomials of degree $\leq n$ with real coefficients. Find the Jordan form for the operator $A = \frac{d^2}{dx^2} : f \in V \rightarrow f'' \in V$. What is the minimal polynomial of $A$?
4.7 The Jordan form

With the spectral theorem and the theory for nilpotent operators, we can now make similarity transformations to associate each square matrix with a canonical form.

**Definition 4.5.** A Jordan block \( J_k(\lambda) \) is a \( k \times k \) matrix of the form

\[
J_k(\lambda) = \lambda I_k + J_k, \quad I_k = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{k \times k}, \quad J_k := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \ddots \end{pmatrix}_{k \times k}
\]

namely, \( I_k \) is the identity matrix of rank \( k \), and \( J_k = (\delta_{i+1,j})_{k \times k} \). When \( k = 1 \), \( J_1 = (0) \).

A Jordan matrix is a direct sum of Jordan blocks:

\[
J = \begin{pmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & J_{k_m}(\lambda_m) \end{pmatrix}
\]

in which the values \( \lambda_i \) need not be distinct.

Given a matrix \( A \in M_n(\mathbb{F}) \), if \( J = P^{-1}AP \) is a Jordan matrix, then \( J \) is also called the Jordan canonical form or Jordan normal form of the matrix \( A \).

We remark that one can require the Jordan blocks to be arranged in a special order so that the Jordan canonical form of a square matrix is unique. We shall not do it here.

**Theorem 4.4. (The Jordan Theorem of Normal Form)**

Any square matrix on a closed field is similar to a Jordan matrix.

For every linear operator from a finite dimensional vector space over a closed field into itself, there exists a basis such that the matrix of the operator under the basis is a Jordan matrix.

We leave the proof as an exercise.

From the process of finding basis for nilpotent operators, we see that the final canonical form depends only on the dimensional of \( \ker(A^j) \), \( j = 1, \ldots, n \), or \( \ker((A - \lambda I)^m) \) on \( E(\lambda) \).

Indeed, we have the following:

**Theorem 4.5.** Let \( V \) be a finite dimensional vector space over a closed field \( \mathbb{F} \) and \( A \) and \( B \) be two linear maps from \( V \) into itself. Then \( A \) is similar to \( B \), i.e., there exists an automorphism \( P \) such that \( B = P^{-1}AP \) if and only if

\[
\dim \left( \ker((\lambda I - A)^m) \right) = \dim \left( \ker((\lambda - B)^m) \right) \quad \forall \lambda \in \mathbb{F}, \quad m \geq 1.
\]

Instead of giving a proof to the Jordan theorem, we provide here the basic steps in finding the Jordan normal form for given a square matrix \( A \) in \( M_n(\mathbb{F}) \).

1. Find the characteristic polynomial \( p(z) = \det(zI - A) \) and all the eigenvalues: \( \lambda_i \) with multiplicity \( m_i \), \( i = 1, \cdots, n \). Then for each \( i \), set \( B = A - \lambda_i I \) and perform the following step.
(2)(i) Find \( N_0, N_1, \ldots, N_k \) by solving iteratively the following
\[
B^{-1}(0) = N_0 \implies \ker(B^1) = N_0, \\
B^{-1}(N_{j-1}) = N_j \oplus N_0 \implies \ker(B^{j+1}) = N_j \oplus \ker(B^j), \quad BN_j \subseteq N_{j-1}.
\]
In finite steps, one finds \( N_{k+1} = \{0\} \). This implies that
\[
\mathbf{E}(\lambda_i) = \cup_{j \geq 1} \ker(B^j) = \ker(B^{k+1}) = N_k \oplus N_{k-1} \oplus \cdots \oplus N_0.
\]
Denote by \( m_i \) the multiplicity of the eigenvalue \( \lambda_i \) and by \( n_j \) the dimension of \( N_j \). Then
\[
0 = n_{k+1} < n_k \leq n_{k-1} \leq \cdots \leq n_0, \\
\dim(\mathbf{E}(\lambda_i)) = m_i = \sum_j \dim(N_j) = n_k + \cdots + n_0.
\]
(ii) Find a basis \( U_k = \{x_1, \ldots, x_{n_k}\} \) for \( N_k \). For each \( j = k-1, k-2, \ldots, 0 \), generate a basis \( U_j \) for \( N_j \) by adding vectors \( x_{n_{j+1}}, \ldots, x_{n_{j-1}} \) to an initial collection \( BU_{j+1} \).
(iii) Combine \( U_0, \ldots, U_k \) and order them to a basis for \( \mathbf{E}(\lambda_i) \) in the following form
\[
P_i = (B^{k_1}x_1, \ldots, x_1, \ldots, k_0x_{n_0}, \ldots, x_{n_0}).
\]
Then \( AP_i = P_iJ^i \) where \( J^i = \text{diag}(J_{k_1}, \ldots, J_{k_0}) \) consists of \( n_0 \) Jordan blocks.
(3) Finally combining all the bases \( P_i \) for \( \mathbf{E}(\lambda_i), i = 1, \ldots, \hat{n} \), we obtain the similarity transform matrix \( P = (P_1 \cdots P_{\hat{n}}) \) and the Jordan form \( J = \text{diag}(J^1, \ldots, J^{\hat{n}}) \).

**Example 4.9.** Find the Jordan form and the associated transformation for the matrix
\[
A = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 \\
2 & 2 & 0 & 1 \\
3 & 3 & 0 & 2
\end{pmatrix}
\]

**Solution.** (1) The characteristic polynomial is \( p(z) = (z-1)^4(z-2) \). Hence, \( \lambda_1 = 2 \) is a simple eigenvalue, and \( \lambda_2 = 1 \) is an eigenvalue of multiplicity 4.
(2) For \( \lambda_1 = 2 \), we find an eigenvector \( \xi_1 = (0 \ 0 \ 0 \ 0 \ 1)^T \).
(3) For \( \lambda_2 = 1 \), solving \( (A - I)x = 0 \) gives us \( \xi_2 = (0 \ 0 \ 1 \ 0 \ 0)^T \), \( \xi_3 = (0 \ 0 \ 0 \ 1 \ 0)^T \), and
\[
N_0 = \ker(A - I) = \text{span}\{\xi_2, \xi_3\}.
\]
(4) Solving \( (A - I)x = c_1\xi_2 + c_2\xi_3 \in N_0 \) gives us \( \xi_4 = (0 \ 1 \ 0 \ 0 \ -3)^T \) and
\[
N_1 = \text{span}\{\xi_4\}, \quad \ker((A - I)^2) = N_1 \oplus N_0, \quad (A - I)\xi_4 = \xi_2 + 2\xi_3.
\]
(5) Solving \( (A - I)x = x_4 \in N_1 \) gives \( \xi_5 = (1 \ -1 \ 0 \ 0 \ -3)^T \) and
\[
N_2 = \text{span}\{\xi_5\}, \quad \ker((A - I)^3) = N_2 \oplus \ker((A - I)^2), \quad (A - I)\xi_5 = \xi_4.
\]
(6) Now we go back. Set \( U_2 = \{\xi_5\}, U_1 = \{(A - I)\xi_5\} = \{\xi_4\}, U_0 = \{(A - I)\xi_4, \xi_3\} \). Hence, we set
\[
\begin{align*}
\mathbf{u}_3 &= \xi_5, \\
\mathbf{u}_2 &= (A - I)\mathbf{u}_3 = \xi_4, \\
\mathbf{u}_1 &= (A - I)^2\mathbf{u}_3 = \xi_2 + 2\xi_4, \quad (A - I)\mathbf{u}_1 = 0, \\
\mathbf{u}_4 &= \xi_3, \quad (A - I)\mathbf{u}_4 = 0 \\
\mathbf{u}_5 &= \xi_1, \quad (A - 2I)\xi_1 = 0
\end{align*}
\]
4.7. THE JORDAN FORM

We see that
\[ A(u_1, u_2, u_3, u_4, u_5) = (u_1 + u_2 + u_3, u_3, u_4, 2u_5). \]

That is, \( AP = PJ \) or \( P^{-1}AP = J \) where
\[
P = \begin{pmatrix}
u_1 & u_2 & u_3 & u_4 & u_5 \\
0 & 1 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{pmatrix}
\]
\[
J = \begin{pmatrix}1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 2\end{pmatrix}
\]

Exercise 28. Prove the Jordan theorem of normal form for square matrices.

Exercise 29. Prove Theorem 4.5. Also state its analogous for matrices.

Exercise 30. Let \( A \in M_5(\mathbb{R}) \). Find possible Jordan forms in the following cases:

(i) \( A \) is idempotent: \( A^2 = A \);
(ii) \( A \) is tripotent: \( A^3 = A \);
(iii) \( A^k = A \) for some \( k \geq 4 \).

Exercise 31. For the following matrices, find \( P \) such that \( P^{-1}AP \) is in a Jordan form:

\[
\begin{pmatrix}1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 0 & 1 & 0 \\
3 & 3 & 0 & 0 & 1\end{pmatrix}, \quad \begin{pmatrix}1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 0 & 1 & 0 \\
3 & 4 & 0 & 0 & 2\end{pmatrix}, \quad \begin{pmatrix}1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 0 & 1 & 0 \\
5 & 5 & 0 & 0 & 2\end{pmatrix}, \quad \begin{pmatrix}1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 5\end{pmatrix}
\]
4.8 Real Jordan Canonical Forms

In addition to the Jordan canonical form, there are several other matrix factorizations that can be useful in various circumstances. Here we consider only a variant of the Jordan form when the matrix has only real entries. (Recall that \( \mathbb{R} \) is not closed.)

When \( A \) is a real matrix, all the non-real eigenvalues must occur in conjugate pairs. Moreover,
\[
\ker((A - \lambda I)^j) = \ker((A - \lambda I)^k) \quad \forall \lambda \in \mathbb{C}, \ k \geq 1.
\]
Hence, the structure of \( A \) on \( \mathbb{E}(\lambda) \) is the same as that on \( \mathbb{E}(\bar{\lambda}) = \mathbb{E}(\lambda) \).

Suppose \( \lambda = \alpha + \omega i \) is an eigenvalue (\( \omega > 0 \)) and \( \{u_1, \ldots, u_k\} \) is a set of (generalized) eigenvectors on which the corresponding map \( f: x \in \mathbb{C}^n \to Ax \in \mathbb{C}^n \) takes a Jordan block form:
\[
(Au_1 \cdots Au_k) = \lambda(u_1 \cdots u_k) + (0 \ u_1 \cdots u_{k-1}) = (u_1 \cdots u_k)J_k(\lambda).
\]
Write \( u_i = u_i^1 + iu_i^2 \). The invariant space \( \text{span}\{u_1, \ldots, u_k\} \oplus \text{span}\{\bar{u}_1, \ldots, \bar{u}_k\} \) has a real basis \( \{u_1^1, u_1^2, \ldots, u_k^1, u_k^2\} \). Under this basis, we can calculate, since \( A \) is real,
\[
(Au_1^1, Au_1^2) = (\alpha u_1^1 - \omega u_1^2, \omega u_1 + \alpha u_1^2) + (u_{i-1}^1, u_{i-1}^2), \quad i = 1, \ldots, k
\]
where \( u_0^1 = u_0^2 = 0 \). In the matrix form, this can be written as
\[
A(u_1^1 u_1^2 \cdots u_k^1 u_k^2) = (u_1^1 u_1^2 \cdots u_k^1 u_k^2)C_k(\alpha, \omega)
\]
where \( C_k(\alpha, \omega) \) is the real Jordan block
\[
C_k(\alpha, \omega) := \begin{pmatrix} C & I & 0 \\ C & I & \vdots & \ddots \\ \vdots & \ddots & \ddots \\ 0 & \cdots & C & I \\ \end{pmatrix}_{2k \times 2k}, \quad C = C(\alpha, \omega) := \begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix}
\]
(4.2)

and \( I \) is the \( 2 \times 2 \) identity matrix.

Converting each pair of complex Jordan blocks into a real Jordan block, we conclude as follows:

**Theorem 4.6.** Each real matrix \( A \in M_n(\mathbb{R}) \) is similar to, via a real similarity transformation \( P^{-1}AP \), a real Jordan canonical form:
\[
\begin{pmatrix} C_{k_1}(\alpha_1, \omega_1) & 0 & \cdots \\ \vdots & C_{k_p}(\alpha_p, \omega_p) & \cdots \\ 0 & \cdots & J_{m_q}(\lambda_q) \end{pmatrix}
\]
where \( \lambda_i, i = 1, \ldots, q \) are real eigenvalues, and \( \alpha_j + \omega_j i, j = 1, \ldots, p \) (\( \omega_j > 0 \)) are non-real eigenvalues.

**Example 4.10.** Find the real Jordan canonical form and the similarity transformation for \( A = \begin{pmatrix} 0 & 10 \\ -1 & 2 \end{pmatrix} \).

**Solution.** The characteristic polynomial is \( p(z) = z^2 - 2z + 10 \) so eigenvalues are \( \lambda = 1 \pm 3i \).
4.8. REAL JORDAN CANONICAL FORMS

The eigenvector for $1 + 3i$ is $x = (1 - 3i \ 1)^T$. Hence we take

$$u_1 = \text{Re}(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad u_2 = \text{Im}(x) = \begin{pmatrix} -3 \\ 0 \end{pmatrix}, \quad P = (u_1 \ u_2) = \begin{pmatrix} 1 & -3 \\ 1 & 0 \end{pmatrix}.$$  

We find that

$$Au_1 = u_1 - 3u_2, \quad Au_2 = 3u_1 + u_2;$$

or

$$AP = PJ, \quad P^{-1}AP = J = C_1(1, 3) = \begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}.$$

Example 4.11. Find the similarity transformation and the real Jordan normal form for

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

Solution. (i) The characteristic polynomial is $p(z) = \det(zI - A) = [(z - 1)^2 + 1]^2$; the eigenvalues are $1 \pm i$ of multiplicity 2.

(ii) For $\lambda = 1 + 1$, solving $(A - \lambda I)x = 0$ gives

$$\ker(A - \lambda I) = \text{span}\{x_1\}, \quad x_1 = (0 \ 0 \ 1 \ i)^T.$$

(iii) Solving $(A - \lambda I)x = x_1$ gives

$$\ker((A - \lambda I)^2) = \text{span}\{x_2, x_1\}, \quad x_2 = (1 \ i \ 0 \ -i)^T, \quad (A - \lambda I)x_2 = x_1,$$

(iv) Extracting the real and imaginary parts, we obtain

$$P = \begin{pmatrix} \text{Re}(x_1) & \text{Im}(x_1) & \text{Re}(x_2) & \text{Im}(x_2) \end{pmatrix}, \quad AP = PJ, \quad P^{-1}AP = J,$$

where

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \quad J = \begin{pmatrix} 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

Exercise 32. Suppose the minimal polynomial of $A \in M_n(\mathbb{R})$ is $p = z^n - z$. Find its real Jordan normal form.

Exercise 33. Find the similarity transformation and the real Jordan normal form for the following matrices

$$\begin{pmatrix} 1 & 5 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 1 & 1 \\ -2 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 2 & 1 \end{pmatrix}$$

Exercise 34. Let $\mathbb{F} = \mathbb{R}$ and $V$ be the set of all polynomials of order $< 5$. Set $A = \frac{d^2}{dx^2} + 2 \frac{d}{dx} + I$. Find the corresponding matrix for $A^{10}$ under the basis $\{e_0, \cdots, e_4\}$ where $e_i(x) = x^i$ for $i = 0, \cdots, 4$. 
