SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Jordan Normal Form

av

Johannes Erixon

2025 - No K24

Jordan Normal Form

Johannes Erixon

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Wushi Goldring

2025

Sammanfattning

Arbetet behandlar Jordan normalform som ett verktyg för att analysera linjära operatorer, särskilt i fall där diagonalisation inte är möjlig. Genom exempel och teori visas hur funktioner som $e^A, \sin(A)$ och $\log(I + A)$ kan beräknas via Jordanformen, och hur denna struktur ger insikt i operatorers algebraiska egenskaper.

Abstract

This thesis explores Jordan Normal Form as a method for analyzing linear operators, especially when diagonalization fails. Through examples and theory, it demonstrates how functions such as $e^A, \sin(A)$ and $\log(I + A)$ can be computed using the Jordan structure, revealing deeper algebraic properties of operators.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Wushi Goldring, whose insights and guidance have been invaluable throughout the course of this work. A big thank you also to all the teachers at the Department of Mathematics, whose pedagogical skills and commitment have inspired me throughout my education. I would also like to thank my parents, Stig and Anna-Karin Erixon, for supporting me and giving me encouragement throughout my work. Finally, I would like to mention Copilot, whose AI-based reasoning and reflections have served as a creative and stimulating conversation partner through many of the most difficult and rewarding parts of the work.

Contents

1 Introduction

Eigenvalues and eigenvectors are central concepts in linear algebra, providing insight into the structure of linear transformations. They are not only of theoretical interest but also have practical significance in the analysis of stability, matrix decompositions, and operator behavior. This thesis explores these ideas with particular focus on diagonalization and Jordan Normal Form, following the presentation and theory as outlined in Axler's *Linear Algebra Done Right*[1].

1.1 Mathematical Motivation

Eigenvalues describe how a linear operator stretches or compresses space along certain directions, while eigenvectors identify those invariant directions. Diagonalization makes these relationships transparent by expressing the operator in a simplified form. However, not all operators are diagonalizable — in such cases, the Jordan Normal Form provides a more general structure by incorporating generalized eigenvectors and nilpotent components.

1.2 Structure of the Thesis

This thesis is divided into five chapters:

- **Chapter 2: Eigenvalues and Eigenvectors** – Definitions and properties of eigenvalues, eigenvectors, and minimal polynomials.
- **Chapter 3: Diagonalization** – Conditions under which a matrix or operator is diagonalizable.
- **Chapter 4: Jordan Normal Form** – Construction of Jordan blocks and generalized eigenvectors, and how Jordan form enables the computation of functions of matrices.
- **Chapter 5: Concluding Reflections** – A final synthesis of the main ideas, highlighting the structural insights provided by Jordan form and its relevance beyond diagonalization.

1.3 Objectives and Scope

The primary goal is to give a self-contained treatment of eigenvalues and matrix decomposition, using Axler (2024) as the sole reference. Special emphasis is placed on Jordan Normal Form as a canonical generalization of diagonalization. Along the way, the thesis aims to develop both theoretical understanding and practical methods for analyzing linear operators.

Transition to Chapter 2. Having outlined the purpose and structure of this thesis, we now turn to the mathematical core: the notion of eigenvalues and eigenvectors. These concepts form the gateway to understanding how linear operators behave — and they serve as the foundational tools for all subsequent decomposition theorems.

2 Eigenvalues and Eigenvectors

Introduction

Chapter 2 establishes the basic language and central concepts of linear mappings, with a particular focus on the spectral properties of operators. The aim is to provide a robust terminological and algebraic platform for the more advanced structures discussed in Chapter 4. By introducing concepts such as eigenvalues, minimal polynomials, and various types of multiplicities, a context is created in which diagonalizability and generalized eigenspaces can be understood in depth.

2.1 Foundational Definitions

In this section, we lay the groundwork for understanding eigenvalues and eigenvectors.

Definition 2.1.1 (Linear operator)**.** *Let $\mathbb{F}$ be a field. Let V be an $\mathbb{F}$-vector space. A* ***linear operator*** *is a map $T : V \to V$ that preserves linearity:*

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

for all $v, w \in V$ and all scalars $a, b \in \mathbb{F}$.

Linear operators form the basis for analyzing transformations within vector spaces. This definition establishes the framework for understanding linear transformations that act within the same vector space. Linear operators are crucial in exploring how vectors transform under mappings, offering insights into structures such as invariant subspaces and eigenvalues.

Definition 2.1.2 (Invariant subspace, Axler (2024, p. 126))**.** *Let $T \in \mathcal{L}(V)$ be a linear operator. A subspace $U \subseteq V$ is said to be* invariant *under T if*

$$T(u) \in U \quad \textit{for all } u \in U.$$

Example 2.1.3 (Based on Axler (2024, p. 133))**.** *For a general operator $T \in \mathcal{L}(V)$, common invariant subspaces include:*

- *The zero subspace $\{0\}$, as $T(0) = 0$.*
- *The entire space V, as $T(u) \in V$ for all $u \in V$.*
- *The null space null T, since $T(u) = 0$ implies $u \in$ null T.*
- *The range range T, as $u = T(v)$ ensures $T(u) \in$ range T.*

Particularly significant are one-dimensional invariant subspaces, which are directly tied to eigenvalues, where the operator T acts as scalar multiplication:

$$T(v) = \lambda v \quad \text{for some scalar } \lambda.$$

These foundational examples highlight how invariant subspaces are structured and pave the way for understanding eigenvalues and eigenvectors in subsequent sections.

Definition 2.1.4 (Eigenvalues and Eigenvectors). *Eigenvalues λ are scalars satisfying $T(v) = \lambda v$ for a nonzero vector v, known as an eigenvector. Eigenvectors form one-dimensional invariant subspaces where T acts as scalar multiplication.*

Example 2.1.5 (Eigenvalues and Eigenvectors. Based on Axler (2024, p. 135)). *Consider $T \in \mathcal{L}(\mathbb{F}^2)$, defined as $T\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} -z \\ w \end{pmatrix}$.*

- *In $\mathbb{R}^2$, T represents a 90° rotation, which yields no eigenvalues.*
- *In $\mathbb{C}^2$, eigenvalues $\lambda = i$ and $\lambda = -i$ exist, with eigenvectors $\begin{pmatrix} u \\ -iu \end{pmatrix}$ and $\begin{pmatrix} u \\ iu \end{pmatrix}$, where $u \neq 0$.*

Definition 2.1.6 (Characteristic Polynomial). *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ be a field, L a finite-dimensional vector space over $\mathbb{F}$, and $T : L \to L$ a linear operator. The characteristic polynomial $P(t)$ of T is defined as:*

$$P(t) = \det(tI - M),$$

where M is the matrix representation of T with respect to a chosen basis of L, I is the identity matrix, and det *denotes the determinant operation.*

The roots of $P(t)$ are the eigenvalues of T, providing a fundamental connection between the operator's algebraic properties and its action on the vector space.

Definition 2.1.7 (Polynomial applied to an operator, Axler (2024, p. 137)). *Let*

$$p(t) = a_0 + a_1 t + \cdots + a_n t^n$$

be a polynomial, and let $T \in \mathcal{L}(V)$. Then:

$$p(T) = a_0 I + a_1 T + \cdots + a_n T^n,$$

where I is the identity operator.

Example 2.1.8. *For example, if $p(t) = t^2 + 2t + 1$, applying $p(T)$ results in:*

$$p(T) = T^2 + 2T + I.$$

This operation links the matrix representation of T to geometric properties such as scaling and direction preservation in invariant subspaces.

Example 2.1.9 (Adapted from Axler (2024, p. 138))**.** *Let $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ be the differentiation operator, defined by $Dq = q'$. Consider the polynomial $p(x) = 7 - 3x + 5x^2$. Applying this polynomial to the operator D gives:*

$$p(D) = 7I - 3D + 5D^2$$

Thus, for any $q \in \mathcal{P}(\mathbb{R})$, we have:

$$(p(D))q = 7q - 3q' + 5q''$$

Before we get to the definition of the minimal polynomial, we define the concepts of algebraic multiplicity and geometric multiplicity.

Definition 2.1.10 (Algebraic multiplicity)**.** *The algebraic multiplicity of an eigenvalue λ is the exponent of the factor $(x - \lambda)$ in the characteristic polynomial.*

Definition 2.1.11 (Geometric multiplicity)**.** *The geometric multiplicity of λ is the dimension of the eigenspace $E(\lambda, T) = null(T - \lambda I)$.*

The characteristic polynomial reveals which eigenvalues an operator has and their algebraic multiplicity, while the geometric multiplicity expresses how many linearly independent eigenvectors belong to each eigenvalue. These multiplicities provide important information about the diagonalizability of the operator: when the geometric multiplicity is equal to the algebraic multiplicity for all eigenvalues, the operator is diagonalizable. The minimal polynomial, on the other hand, captures more than just eigenvalues it specifies the smallest degree polynomial that annihilates the operator. The exponent of each factor $(x - \lambda)$ in the minimal polynomial determines the length of the longest chains of generalized eigenvectors.

Definition 2.1.12 (Minimal Polynomial, Axler (2024, p. 145))**.** *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the* minimal polynomial *of T is the unique monic polynomial $p \in \mathcal{P}(\mathbb{F})$ of smallest degree such that*

$$p(T) = 0.$$

The minimal polynomial reveals the operator's eigenvalues and provides a pathway to understanding its invariant subspaces and nilpotent structures.

Example 2.1.13 (Finding the Minimal Polynomial)**.** *Let $T \in \mathcal{L}(V)$, and consider its matrix representation relative to the standard basis $\{e_1, e_2, e_3, e_4, e_5\}$:*

$$M(T) = \begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The action of T on the basis vectors is:

$$T(e_1) = e_2, \quad T(e_2) = e_3, \quad T(e_3) = e_4, \quad T(e_4) = e_5, \quad T(e_5) = -3e_1 + 6e_2.$$

Through successive applications of T, we compute:

$$T^2(e_1) = e_3, \quad T^3(e_1) = e_4, \quad T^4(e_1) = e_5, \quad T^5(e_1) = -3e_1 + 6e_2.$$

From this, the minimal polynomial of T is deduced to be:

$$p_T(x) = 3 - 6x + x^5.$$

In this example, we could construct the minimal polynomial in the same way as the characteristic one, but with adjusted exponents for factors where the geometric multiplicity is less than the algebraic one. This way, it becomes clear how the two polynomials relate to each other and to the diagonalizability of the operator.

Significance: Minimal Polynomial

The minimal polynomial plays a critical role in analyzing T's invariant subspaces, eigenvalues, and subsequent transformations into Jordan Normal Form. It bridges the gap between algebraic theory and practical matrix analysis.

Summary. This section established the foundational language of linear operators, invariant subspaces, and eigenstructures. Eigenvalues and eigenvectors provide insight into how operators act along specific directions, while the characteristic and minimal polynomials encode spectral and algebraic information. Algebraic and geometric multiplicities play a key role in understanding diagonalizability, and the minimal polynomial provides a bridge to more refined structures such as generalized eigenspaces and Jordan Normal Form.

2.2 Theorems and Examples

In this section, we formalize and explore key results concerning linear operators and eigenvalues. These theorems establish the mathematical framework for analyzing operators and serve as the foundation for practical computations involving eigenvalues and their relationships to invariant subspaces.

Theorem 2.2.1. *Let $T \in \mathcal{L}(V)$ be a linear operator on a finite-dimensional vector space V. Then the characteristic polynomial of T is independent of the choice of basis.*

Proof. Let A be a matrix representation of T in some basis, and let $A' = P^{-1}AP$ be the matrix of T in another basis, with P invertible. Then:

$$\det(xI - A') = \det(xI - P^{-1}AP) = \det\left(P^{-1}(xI - A)P\right)$$

By the multiplicative property of determinants:

$$= \det(P^{-1}) \cdot \det(xI - A) \cdot \det(P) = \det(xI - A) \cdot \underbrace{\det(P^{-1})\det(P)}_{=1}$$

Thus, $\det(xI - A') = \det(xI - A)$, and the characteristic polynomial is basis-independent. □

Theorem 2.2.2. *Let $T \in \mathcal{L}(V)$ be a linear operator on a finite-dimensional vector space V, and let $p_T(x) = \det(xI - A)$ be the characteristic polynomial of a matrix representation A of T. Then:*

A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if it is a root of the characteristic polynomial $p_T(x)$.

Proof. ($\Rightarrow$) Suppose that λ is an eigenvalue. Then $A - \lambda I$ is non-invertible, which occurs precisely when λ is a root of $\det(xI - A)$. Then there exists a non-zero vector $v \in V$ such that: $T(v) = \lambda v \Rightarrow (T - \lambda I)v = 0$. This means that $T - \lambda I$ is not injective. In a basis, the operator has matrix A, so:

$$(A - \lambda I)v = 0 \text{ for } v \neq 0.$$

Thus λ is a zero of the characteristic polynomial.
($\Leftarrow$)Now assume that λ is a zero of $p_T(x) = \det(xI - A)$. Then

$$\det(\lambda I - A) = 0.$$

Thus $\lambda I - A$ is non-invertible $\Rightarrow (A - \lambda I)v = 0$ has a non-trivial solution $\Rightarrow$ there is a $v \neq 0$ with:

$$T(v) = \lambda v.$$

Thus λ is an eigenvalue.

$\square$

The characteristic polynomial is a central algebraic tool for identifying eigenvalues, enabling a deeper analysis of operator properties. Following this, the equivalent conditions for eigenvalues provide a comprehensive characterization of their algebraic and operator-theoretic implications.

Theorem 2.2.3 (Equivalent Conditions for Eigenvalues, cf. Axler (2024, p. 135))**.** *Let $T \in \mathcal{L}(V)$ be a linear operator on a finite-dimensional vector space V, and let $\lambda \in \mathbb{F}$. The following conditions are equivalent:*

1. *λ is an eigenvalue of T.*
2. *$T - \lambda I$ is not injective.*
3. *$T - \lambda I$ is not surjective.*
4. *$T - \lambda I$ is not bijective.*

Proof.

$1 \Leftrightarrow 2$**:** By definition, λ is an eigenvalue of T if there exists $v \neq 0$ such that

$$T(v) = \lambda v \quad \Leftrightarrow \quad (T - \lambda I)v = 0.$$

Hence, $T - \lambda I$ is not injective.

$2 \Leftrightarrow 3 \Leftrightarrow 4$**:** Since V is finite-dimensional, injectivity, surjectivity, and bijectivity are equivalent conditions for linear maps. Thus, if one of them fails, all fail.

$\square$

The equivalence between injectivity and surjectivity for operators on finite-dimensional spaces relies on the following fundamental result. Having established the theoretical conditions for eigenvalues, we now apply them to specific cases to illustrate their practical computation and connection to matrix representations.

Example 2.2.4 (Complex versus Real Eigenvalues, cf. Axler (2024),p. 135 Example 5.9)**.** *Consider the linear operator $T \in \mathcal{L}(\mathbb{F}^2)$ defined by*

$$T\begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} -z \\ w \end{pmatrix}.$$

We will examine the eigenvalues of T when acting on $\mathbb{R}^2$ versus on $\mathbb{C}^2$.

Matrix Representation: *In the standard basis $\{(1,0),(0,1)\}$, the matrix of T is*

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

Characteristic Polynomial:

$$\det(xI - A) = \det\begin{pmatrix} x & 1 \\ -1 & x \end{pmatrix} = x^2 + 1.$$

This polynomial has no real roots*, so T has no real eigenvalues when considered as an operator on $\mathbb{R}^2$. The same holds for eigenvectors: no nonzero real vector satisfies $T(v) = \lambda v$ for real λ.*

Complex Case: *Over $\mathbb{C}^2$, the eigenvalues are the roots of $x^2 + 1$, i.e.,*

$$\lambda = i \quad \textit{and} \quad \lambda = -i.$$

Let's find eigenvectors corresponding to $\lambda = i$. We solve

$$(A - iI)v = 0 \quad \Rightarrow \quad \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}\begin{pmatrix} w \\ z \end{pmatrix} = 0.$$

Solving the first row: $-iw - z = 0 \Rightarrow z = -iw$. So an eigenvector is

$$v = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

In other words, $T\begin{pmatrix} 1 \\ -i \end{pmatrix} = i\begin{pmatrix} 1 \\ -i \end{pmatrix}$.

Conclusion: *This example shows that the existence of eigenvalues (and diagonalizability) can depend on the choice of field. The operator T has no eigenvalues over $\mathbb{R}$, but becomes fully diagonalizable over $\mathbb{C}$.*

Theorem 2.2.5 (Linear Independence of Eigenvectors Axler (2024) p. 136)**.** *If $T \in \mathcal{L}(V)$, any eigenvectors of T corresponding to distinct eigenvalues are linearly independent.*

Proof. We prove the result by induction on m, the number of eigenvectors.

For $m = 1$, the set $\{v_1\}$ is linearly independent since $v_1 \neq 0$ (by definition of eigenvector).

Assume now that $\{v_1, \ldots, v_{m-1}\}$ is linearly independent for some $m \geq 2$. Suppose

$$a_1 v_1 + \cdots + a_m v_m = 0,$$

and we want to show $a_1 = \cdots = a_m = 0$. Apply T to both sides:

$$a_1 \lambda_1 v_1 + \cdots + a_m \lambda_m v_m = 0.$$

Subtracting λ_m times the original equation gives:

$$a_1(\lambda_1 - \lambda_m)v_1 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

By the induction hypothesis, the set $\{v_1, \ldots, v_{m-1}\}$ is linearly independent, and since the eigenvalues are distinct, each $\lambda_j - \lambda_m \neq 0$. Hence:

$$a_1 = \cdots = a_{m-1} = 0.$$

Returning to the original equation, this gives $a_m v_m = 0$, and since $v_m \neq 0$, we have $a_m = 0$. Thus all $a_j = 0$, and the set is linearly independent. □

Linear independence follows directly from the distinctness of eigenvalues, as any dependence would contradict this property.

Significance

The linear independence of eigenvectors ensures that operators with distinct eigenvalues admit a complete eigenbasis. This is a critical step in the diagonalization process, as it enables the transformation of the operator into a diagonal form.

Theorem 2.2.6 (Invariance of Null Space and Range, cf. Axler (2024, p.139, Thm. 5.18))**.** *Let $T \in \mathcal{L}(V)$, and let $p \in \mathcal{P}(\mathbb{F})$ be a polynomial with coefficients in the field $\mathbb{F}$. Then both the null space and the range of $p(T)$ are invariant under T.*

Proof. Suppose $u \in \operatorname{null} p(T)$. Then $p(T)u = 0$, so

$$p(T)(Tu) = T(p(T)u) = T(0) = 0.$$

Thus $Tu \in \operatorname{null} p(T)$, showing that the null space is invariant under T.

Now suppose $u \in \operatorname{range} p(T)$, so there exists $v \in V$ such that $u = p(T)v$. Then

$$Tu = T(p(T)v) = p(T)(Tv),$$

which shows that $Tu \in \operatorname{range} p(T)$, and hence the range is invariant under T. $\square$

Thus, polynomial operators preserve the null space and range of T, ensuring structural consistency and invariance.

By the Fundamental Theorem of Algebra, any linear operator on a finite-dimensional complex vector space has at least one eigenvalue.

While eigenvalues provide critical insights into finite-dimensional operators, the minimal polynomial offers a broader tool for analyzing operator properties. The following theorem formalizes the existence and uniqueness of the minimal polynomial in finite-dimensional spaces.

Theorem 2.2.7 (Minimal Polynomial, cf. Axler (2024, Thm. 5.23))**.** *Let $T \in \mathcal{L}(V)$, where V is a finite-dimensional vector space. Then there exists a unique monic polynomial $p \in \mathbb{F}[x]$ of smallest possible degree such that*

$$p(T) = 0.$$

This polynomial is called the minimal polynomial *of T, and it has degree at most* $\dim V$.

Proof. We proceed by induction on $\dim V$.

If $\dim V = 0$, then $V = \{0\}$, so every operator on V is the zero operator and thus satisfies $p(T) = 0$ for any polynomial p. In this case, we can define the minimal polynomial to be $p(x) = 1$.

Now assume that $\dim V > 0$, and that the statement holds for all operators on vector spaces of lower dimension. Pick a nonzero vector $v \in V$. Since V is finite-dimensional, the list

$$v,\ Tv,\ T^2v,\ \ldots,\ T^m v$$

must be linearly dependent for some $m \le \dim V$. This implies that there exists a monic polynomial q of degree m such that $q(T)v = 0$.

Because T is linear, it follows that $q(T)(T^k v) = T^k q(T)v = 0$ for every

$k \geq 0$, so the subspace generated by $\{T^k v\}$ lies in the null space of $q(T)$. Thus, $q(T)$ sends all of V into a proper subspace $W := \text{range}(q(T))$, which is invariant under T. Moreover, $\dim W < \dim V$.

By the induction hypothesis, the restriction $T|_W \in \mathcal{L}(W)$ has a minimal polynomial r. Then the product $p := r \cdot q$ satisfies $p(T) = 0$. Among all such polynomials that annihilate T, we define the minimal polynomial as the unique monic polynomial of smallest degree.

To prove uniqueness, suppose p and s are two monic polynomials of minimal degree such that $p(T) = s(T) = 0$. Then $(p - s)(T) = 0$, and $p - s$ has degree strictly less than $\deg p$ unless $p = s$, contradicting the minimality unless they are equal. Hence, the minimal polynomial is unique. $\square$

Having established the existence and uniqueness of the minimal polynomial, we now explore its role in revealing the spectral structure of the operator, including eigenvalues and their multiplicities.

The Role of the Minimal Polynomial

The minimal polynomial provides critical insights into an operator's spectrum, revealing eigenvalues and their multiplicities. It not only encapsulates the algebraic properties of an operator but also lays the groundwork for subsequent theorems connecting it to eigenspace decomposition and Jordan Normal Form.

Theorem 2.2.8 (Eigenvalues Are Zeros of the Minimal Polynomial, cf. Axler (2024, p. 146))**.** *Let $T \in \mathcal{L}(V)$ where V is finite-dimensional. Then:*

1. *The eigenvalues of T are precisely the zeros of its minimal polynomial $p_T(z)$.*
2. *If $\mathbb{F} = \mathbb{C}$, then $p_T(z)$ factors as*

$$p_T(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m),$$

where $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ are the eigenvalues of T, possibly repeated.

Proof.

1. If λ is an eigenvalue of T, then $T - \lambda I$ is not injective, hence not invertible, so $p_T(T)$ must vanish at λ. Conversely, if $p_T(\lambda) = 0$, then $T - \lambda I$ is not invertible, so λ is an eigenvalue.

2. Over $\mathbb{C}$, the Fundamental Theorem of Algebra ensures that any polynomial factors into linear terms. Since p_T is monic and has the eigenvalues as zeros, the factorization follows.

□

Having established the connection between eigenvalues and the minimal polynomial, we now examine a specific operator to illustrate these principles in practice.

Remark (Factorization of the Minimal Polynomial)**.** *Over $\mathbb{C}$, the Fundamental Theorem of Algebra ensures that the minimal polynomial $p(z)$ factors into linear terms:*

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m),$$

where the λ_i are eigenvalues of T, with repetitions corresponding to algebraic multiplicity.

The minimal polynomial thus provides a complete list of the eigenvalues — including algebraic multiplicities — when $\mathbb{F} = \mathbb{C}$. We now use this perspective to deduce a fundamental property of operators over real vector spaces.

The decomposition into invariant subspaces reveals the intrinsic structure of the operator. We now apply this understanding to analyze operators on odd-dimensional spaces.

Theorem 2.2.9 (Real Eigenvalue in Odd Dimension, cf. Axler (2024, p. 150))**.** *Let V be a finite-dimensional real vector space with $\dim V$ odd. Then every linear operator $T \in \mathcal{L}(V)$ has at least one real eigenvalue.*

Sketch of proof. The characteristic polynomial of T has odd degree with real coefficients, and thus must have at least one real root. Hence, T has a real eigenvalue. □

From Algebraic Structure to Triangular Representations

Having established the foundational results concerning eigenvalues, minimal polynomials, and their implications for diagonalizability, we now turn to a new perspective: how the structure of a linear operator is reflected in its matrix representation. The remainder of this chapter investigates the role of triangular forms, particularly upper-triangular matrices, and connects these representations to the operator's algebraic properties.

Example 2.2.10 (Matrix Representation of an Operator). *Consider* $T \in \mathcal{L}(F^3)$ *defined by* $T(x,y,z) = (2x+y, 5y+3z, 8z)$. *The matrix representation of* T *with respect to the standard basis is:*

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

Theorem 2.2.11 (Triangular Matrices Satisfy Their Characteristic Polynomial, cf. Axler (2024, p. 156)). *Suppose* $T \in \mathcal{L}(V)$ *has an upper-triangular matrix with diagonal entries* $\lambda_1, \ldots, \lambda_n$ *in some basis of* V. *Then*

$$(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I) = 0.$$

Proof. Let A be the upper-triangular matrix of T in some basis. Then the matrix

$$(A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_n I)$$

is also upper-triangular, with zero diagonal (since each diagonal entry of A is subtracted exactly once). An upper-triangular matrix with zeros on the diagonal is nilpotent — so there exists k such that this matrix to the power k is the zero matrix.

Hence,

$$(T - \lambda_1 I) \cdots (T - \lambda_n I)$$

acts as a nilpotent operator. But since the operator is already represented by an upper-triangular nilpotent matrix, we must have:

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0.$$

□

Building on the operator equation, we now formalize the connection between upper-triangular matrices and their eigenvalues.

Theorem 2.2.12 (Eigenvalues from Upper-Triangular Matrices, cf. Axler (2024, p. 157)). *Suppose* $T \in \mathcal{L}(V)$ *has an upper-triangular matrix with respect to some basis, and let* $\lambda_1, \ldots, \lambda_n$ *denote the diagonal entries. Then the eigenvalues of* T *are exactly* $\lambda_1, \ldots, \lambda_n$, *counted with multiplicity.*

Proof. Let A be the upper-triangular matrix of T in some basis. The characteristic polynomial of A is

$$\det(xI - A) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n),$$

since the determinant of an upper-triangular matrix equals the product of its diagonal entries (after shift by x). Therefore, the eigenvalues of T are precisely the diagonal entries $\lambda_1, \ldots, \lambda_n$. □

To illustrate these principles, we examine a concrete example where eigenvalues are derived directly from the diagonal entries.

Example 2.2.13 (Eigenvalues of Upper-Triangular Matrix)**.** *Consider $T \in L(\mathbb{R}^3)$ with matrix:*

$$M(T) = \begin{pmatrix} 2 & 1 & 3 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{pmatrix}.$$

The eigenvalues are $\lambda = 2, 3, 5$, derived directly from the diagonal entries.

This example demonstrates the influence of field selection on matrix properties, paving the way for a deeper analysis of upper-triangular forms in different algebraic settings.

Impact of Field F on Upper-Triangular Form

Let $F = \mathbb{Q}$ or $F = \mathbb{R}$, and consider the polynomials $(x^2 + 3x - 3)^2$ and $(x^2 - 3x + 2)^2$. These forms illustrate how field selection alters upper-triangular matrix representations, enabling more refined algebraic analysis.

Summary. This section developed the fundamental theory connecting eigenvalues to linear operators. Key results include equivalent definitions of eigenvalues, the structure of eigenspaces, and the behavior of polynomial functions applied to operators. The characteristic and minimal polynomials serve as algebraic tools for detecting eigenvalues, and theorems such as rank-nullity and the linear independence of eigenvectors lay the groundwork for understanding diagonalizability and beyond.

Transition to Chapter 3. With a firm grasp of eigenvalues, eigenspaces, and operator structure, we are now ready to examine how — and when — a linear operator can be simplified via diagonalization. This next chapter brings together algebraic and geometric insights to characterize when an operator admits a diagonal form.

3 Diagonalization

Introduction

Chapter 3 deepens the understanding of diagonalizability and shows how linear operators can be expressed in a basis where their action is maximally transparent — by scaling via eigenvalues. This chapter serves as a theoretical bridge between the algebraic concepts of Chapter 2 and the generalized structure developed in Chapter 4. The emphasis is on the criteria and constraints that determine whether an operator is diagonalizable, and thus also on the need for a more general canonical form: the Jordan Normal Form.

3.1 Definition and Conditions for Diagonalization

A diagonal matrix is a square matrix where all non-diagonal entries are zero. For example:

Example 3.1.1 (Diagonal matrix and direct application of f)**.** *Consider the diagonal matrix*

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

This matrix represents a linear operator $T \in \mathcal{L}(\mathbb{F}^3)$ already in diagonal form. Its eigenvalues are $\lambda = 2,\ 5,\ 7$, and the corresponding eigenspaces are:

- $\lambda = 2: \quad E(2,T) = span\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right\},$
- $\lambda = 5: \quad E(5,T) = span\left\{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right\},$
- $\lambda = 7: \quad E(7,T) = span\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}.$

Let f be defined on the spectrum $\{2, 5, 7\}$, and define

$$f(D) = \begin{pmatrix} f(2) & 0 & 0 \\ 0 & f(5) & 0 \\ 0 & 0 & f(7) \end{pmatrix}.$$

which simplifies many operator computations.

This illustrates the power of diagonalization: it reduces the action of a linear operator to scalar-level operations, enabling efficient computation and structural insight. The remainder of this chapter explores the conditions under which such diagonalization is possible.

Example 3.1.2 (Diagonalization and application of f via similarity)**.** *Let*

$$A = \begin{pmatrix} 47 & -15 & 5 \\ 102 & -31 & 12 \\ -54 & 21 & -2 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 5 & 6 \\ -3 & 15 & 2 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

Then

$$A = PDP^{-1},$$

so A is diagonalizable with the same eigenvalues as in Example 3.1.1. Let f be any function defined at $\lambda = 2,\ 5,\ 7$. Then:

$$f(A) = Pf(D)P^{-1} = P \begin{pmatrix} f(2) & 0 & 0 \\ 0 & f(5) & 0 \\ 0 & 0 & f(7) \end{pmatrix} P^{-1},$$

which illustrates that functions of A can be computed via similarity to a diagonal matrix. This diagonalization will be revisited and verified step-by-step in Section 3.3.

Remark (Defining $f(A)$ via diagonalization)**.** *Let f act on the diagonal matrix D. This yields $f(D)$, where the function f is applied pointwise to each diagonal entry — i.e., to each eigenvalue of A. If the matrix representation of the operator T is $A = PDP^{-1}$, with P consisting of eigenvectors of T, then it follows directly that*

$$f(A) = Pf(D)P^{-1}$$

provided that f is defined on the spectrum of A.

Remark (Analytic case)**.** *If the function f is analytic around the eigenvalues of A, then we can also define $f(A)$ via its Taylor series, which justifies the formula $f(A) = Pf(D)P^{-1}$ through term-wise evaluation:*

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = P \left(\sum_{k=0}^{\infty} c_k D^k \right) P^{-1} = Pf(D)P^{-1}.$$

Summary. Diagonalization expresses a linear operator in a basis where it acts by simple scaling — enabling computations like $f(A)$ to reduce to evaluating f on eigenvalues. This section introduced diagonal matrices and showed how similarity transformations make non-diagonal matrices just as computationally accessible, provided they are diagonalizable.

The fact that we can apply a function to a matrix by applying it entrywise to the diagonal — and that this principle also works for diagonalizable matrices — will be formalized and justified later in this chapter. It underpins many applications of diagonalization.

3.2 Fundamental Theorem for Linear Operators

Theorem 3.2.1 (Fundamental Theorem of Linear Maps, cf. Axler (2024, Thm. 3.21, p. 62))**.** *Let $T \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional vector spaces. Then*

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof. Let $\{u_1, \ldots, u_m\}$ be a basis for $\operatorname{null} T$. Extend this to a basis $\{u_1, \ldots, u_m, v_1, \ldots, v_n\}$ for V. Then $\dim V = m + n$.

We claim that $\{Tv_1, \ldots, Tv_n\}$ is a basis for $\operatorname{range} T$. First, it spans: for any $w \in V$, write

$$w = \sum a_i u_i + \sum b_j v_j,$$

so

$$Tw = \sum b_j T v_j,$$

since $Tu_i = 0$. Thus, $Tv_1, \ldots, Tv_n$ spans $\operatorname{range} T$.

To show linear independence, suppose $\sum c_j T v_j = 0$. Then

$$T\left(\sum c_j v_j\right) = 0,$$

so $\sum c_j v_j \in \operatorname{null} T$, and hence expressible as a linear combination of the u_i. But then

$$\sum c_j v_j - \sum d_i u_i = 0,$$

which is a linear combination of the extended basis. Thus all coefficients vanish, so $c_j = 0$. Hence, $Tv_1, \ldots, Tv_n$ is linearly independent.

Therefore, $\dim \operatorname{range} T = n$, and

$$\dim V = m + n = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

$\square$

Significance:
This decomposition simplifies the analysis of linear maps, particularly when studying invariant subspaces and operator diagonalization. It provides a clear structure for analyzing how operators act on finite-dimensional spaces.

This decomposition builds upon the rank–nullity relation (cf. Axler (2024), Theorem 3.21), which guarantees that the dimensions of the kernel and range of a linear map add up to the dimension of the domain. It forms the structural foundation for many subsequent results in this chapter.

Even outside the diagonalizable case, this theorem plays a crucial role. It governs how the kernel and the range interact during iterations of the operator $T - \lambda I$, providing a structural basis for generalized eigenspaces and Jordan blocks. We will return to this in chapter 4.

Example 3.2.2 (Decomposition of $\mathbb{R}^3$). *Consider $f : \mathbb{R}^3 \to \mathbb{R}^3$ with matrix representation:*

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

Here, the kernel and range of f are:

- $\ker(f) = \mathit{span}\{(0, 0, 1)\}$,
- $\mathit{range}(f) = \mathit{span}\{(1, 0, 0), (0, 1, 0)\}$.

This example demonstrates the decomposition $\mathbb{R}^3 = \ker(f) \oplus \operatorname{range}(f)$, illustrating the theorem's application in linear maps.

Connection to Diagonalization:
This decomposition is particularly useful in analyzing conditions for diagonalization, as it reveals invariant subspaces inherent to the operator. In cases where the eigenspaces fail to span the full space, the operator admits a Jordan Normal Form — a framework explored in Chapter 4.

Summary. The Fundamental Theorem of Linear Maps shows that any linear operator on a finite-dimensional space can be decomposed into a kernel–image structure: $V = \ker(T) \oplus V_1$, with $T|_{V_1}$ an isomorphism. This leads to a block matrix form that clarifies how the operator acts. Such decompositions are foundational for analyzing diagonalizability and constructing canonical forms such as the Jordan Normal Form.

3.3 Diagonalization of Operators

We return to the matrix introduced in example 3.1.2 in Section 3.1 to examine whether its diagonalization satisfies the general procedure. We consider the following matrix

$$A = \begin{pmatrix} 47 & -15 & 5 \\ 102 & -31 & 12 \\ -54 & 21 & -2 \end{pmatrix}$$

which is associated with the diagonal matrix

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix}$$

via an expression of the form $A = PDP^{-1}$.

Let us now investigate whether this diagonalization aligns with the general procedure for diagonalizing a matrix: computing the characteristic polynomial, finding the eigenvalues and corresponding eigenspaces, and constructing a basis of eigenvectors to form the change-of-basis matrix P. Before proceeding with the example, we briefly summarize the general procedure for diagonalizing a matrix $A \in \mathbb{R}^{n\times n}$:

1. Compute the characteristic polynomial $\det(\lambda I - A)$.
2. Solve for the eigenvalues λ.
3. For each eigenvalue, compute the corresponding eigenspace.
4. If A has n linearly independent eigenvectors, form the matrix P whose columns are these vectors.
5. Construct the diagonal matrix D with the corresponding eigenvalues on the diagonal.
6. Verify that $A = PDP^{-1}$.

We now apply this method to the matrix from Section 3.1 to check whether the claimed diagonalization indeed holds. We begin by computing the characteristic polynomial of the matrix

$$A = \begin{pmatrix} 47 & -15 & 5 \\ 102 & -31 & 12 \\ -54 & 21 & -2 \end{pmatrix}.$$

We subtract λI and compute the determinant:

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 47 & 15 & -5 \\ -102 & \lambda + 31 & -12 \\ 54 & -21 & \lambda + 2 \end{vmatrix}.$$

Expanding this determinant yields the characteristic polynomial:

$$p_A(\lambda) = \lambda^3 - 14\lambda^2 - 59\lambda + 70 = (\lambda - 2)(\lambda - 5)(\lambda - 7).$$

To find the eigenvalues of A, we solve the equation

$$p_A(\lambda) = 0 \quad \text{that is,} \quad (\lambda - 2)(\lambda - 5)(\lambda - 7) = 0.$$

Solving the characteristic equation above yields the eigenvalues

$$\lambda = 2, \quad \lambda = 5, \quad \lambda = 7.$$

Now we examine the system of equations given by $Av = \lambda v$. If we let $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, then the equation becomes

$$(A - \lambda I) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

For each eigenvalue λ, we solve the corresponding homogeneous system:

(i) For $\lambda = 2$, solve $(A - 2I)v = 0$.

(ii) For $\lambda = 5$, solve $(A - 5I)v = 0$.

(iii) For $\lambda = 7$, solve $(A - 7I)v = 0$.

In each case, we compute the null space of the matrix $A - \lambda I$, which gives the eigenspace $E_\lambda = \ker(A - \lambda I)$. Each nonzero solution vector forms an eigenvector corresponding to λ. These vectors are exactly the ones we already encountered in Section 3.1, namely

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 5 \\ 15 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix}..$$

This gives us the basis exchange matrix

$$P = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 5 & 6 \\ -3 & 15 & 2 \end{pmatrix}.$$

If we now build a diagonal matrix D consisting of eigenvalues along the diagonal with corresponding order as the respective eigenvector in the matrix P, then the diagonal matrix takes the form:

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

Now all that remains is to examine the equality $A = PDP^{-1}$. We find that

$$A = \begin{pmatrix} 47 & -15 & 5 \\ 102 & -31 & 12 \\ -54 & 21 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 5 & 6 \\ -3 & 15 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 2 & 5 & 6 \\ -3 & 15 & 2 \end{pmatrix}^{-1}.$$

This was a simple illustration of an expected result.

Let us now examine the following matrix:

$$A = \begin{pmatrix} -6 & 5 & 10 \\ -2 & 5 & 2 \\ -6 & 3 & 10 \end{pmatrix}.$$

If we now diagonalize this matrix in the same way as the previous one, it will turn out that we arrive at the following result:

$$A = PDP^{-1} \Rightarrow \begin{pmatrix} -6 & 5 & 10 \\ -2 & 5 & 2 \\ -6 & 3 & 10 \end{pmatrix} = \begin{pmatrix} 5 & 4 & 2 \\ 1 & 0 & 2 \\ 3 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 5 & 4 & 2 \\ 1 & 0 & 2 \\ 3 & 4 & 1 \end{pmatrix}^{-1}.$$

The characteristic polynomial is $(\lambda - 4)^2(\lambda - 1)$, which indicates that $\lambda = 4$ is a repeated eigenvalue. Since the eigenspace for $\lambda = 4$ has full dimension two, the operator is still diagonalizable.

According to Theorem 2.2.7, there exists a unique monic minimal polynomial p_T of lowest possible degree such that $p_T(T) = 0$. By Theorem 2.2.8, the minimal polynomial must have exactly the eigenvalues of T as its roots. Therefore, the minimal polynomial is

$$p_T(\lambda) = (\lambda - 4)(\lambda - 1),$$

reflecting the fact that the geometric multiplicity of $\lambda = 4$ equals its algebraic multiplicity.

If we now examine the following matrix instead

$$A = \begin{pmatrix} 14 & -3 & -10 \\ 6 & 1 & -6 \\ 10 & -3 & -6 \end{pmatrix},$$

then it will turn out that this algorithm will no longer work. Rather than compute this decomposition explicitly, we display it directly below.

$$\begin{pmatrix} 14 & -3 & -10 \\ 6 & 1 & -6 \\ 10 & -3 & -6 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 2 \\ 1 & 0 & 2 \\ 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 4 & 2 \\ 1 & 0 & 2 \\ 1 & 4 & 1 \end{pmatrix}^{-1}.$$

We will return to this case in Chapter 4.

Summary. This section applied the diagonalization procedure to concrete matrices, illustrating how the presence and multiplicity of eigenvalues determine whether diagonalization is possible. Operators with a full set of linearly independent eigenvectors can be diagonalized via $A = PDP^{-1}$, but when geometric multiplicities fall short, this decomposition fails. Such cases motivate the Jordan Normal Form, which generalizes diagonalization.

3.4 Eigenspaces and Direct Sums

Diagonalization relies on the ability to decompose a vector space into simpler, invariant components under the action of a linear operator. Eigenspaces provide precisely such components: subspaces where the operator acts by simple scalar multiplication. If these eigenspaces together span the entire space, then the operator can be represented as a diagonal matrix.

In this section, we formalize the notion of an eigenspace and establish the conditions under which a collection of eigenspaces forms a direct sum decomposition of the domain. This provides the structural foundation for understanding when and why a linear operator is diagonalizable. We start with a definition of what we mean by an eigenspace

Definition 3.4.1 (Eigenspace, cf. Axler (2024), Def. 5.52, p. 164)**.** *For $T \in \mathcal{L}(V)$ and eigenvalue λ, the eigenspace $E(\lambda, T)$ is defined as*

$$E(\lambda, T) = \textit{null}(T - \lambda I) = \{v \in V : Tv = \lambda v\}.$$

The eigenspace corresponding to an eigenvalue λ consists of all vectors $v \in V$ such that $(T - \lambda I)v = 0$. This means that $T - \lambda I$ is not invertible — in fact, the existence of a nontrivial solution to this equation is equivalent to the matrix $T - \lambda I$ being singular. Thus, the eigenspace captures precisely those directions in V along which the operator T acts like scalar multiplication by λ. Before we start with a theorem that is central to the context, we recall the definition of direct sum

Definition 3.4.2 (Direct Sum, cf. Axler (2024), Def. 1.41, p. 20)**.** *Let $V_1, \ldots, V_m$ be subspaces of a vector space V. We say that*

$$V = V_1 \oplus \cdots \oplus V_m$$

if every vector $v \in V$ can be written uniquely as

$$v = v_1 + \cdots + v_m, \quad \textit{where } v_k \in V_k.$$

In this case, we say that the sum is direct*, and use the symbol $\oplus$ to indicate that the decomposition is unique.*

Theorem 3.4.3 (Direct Sum of Eigenspaces)**.** *Let $T \in \mathcal{L}(V)$ and $\lambda_1, \ldots, \lambda_m$ be distinct eigenvalues. Then:*

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

if and only if the sum of eigenspaces equals V.

Remark. *A related result appears in Axler (2024), Theorem 5.54, p.164, which states that the sum of eigenspaces corresponding to distinct eigenvalues is always a direct sum — even if it does not span the entire space. Our theorem strengthens this by characterizing when the sum equals V, thus yielding a full decomposition.*

Proof. **(⇒) Sufficiency:** Assume that

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

By definition of a direct sum, each vector $v \in V$ can be written uniquely as

$$v = v_1 + \cdots + v_m, \quad \text{where } v_i \in E(\lambda_i, T).$$

Since each $E(\lambda_i, T) = \ker(T - \lambda_i I)$, it follows that $T(v_i) = \lambda_i v_i$. Thus, applying T to w gives:

$$T(v) = T(v_1 + \cdots + v_m) = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

This demonstrates that T acts diagonally with respect to the decomposition, and since the sum of the eigenspaces spans V, the decomposition is complete.

(⇐) Necessity: Now assume that

$$V = E(\lambda_1, T) + \cdots + E(\lambda_m, T),$$

i.e., the sum of the eigenspaces spans V. To show that the sum is direct, we must prove that the intersection of any combination of eigenspaces is trivial.

Suppose $v \in E(\lambda_i, T) \cap E(\lambda_j, T)$ for some $i \neq j$. Then:

$$T(v) = \lambda_i v = \lambda_j v.$$

Since the eigenvalues are distinct, this implies $(\lambda_i - \lambda_j)v = 0$, and thus $v = 0$. It follows that the pairwise intersections are trivial. By induction, the full intersection is trivial:

$$E(\lambda_1, T) \cap \cdots \cap E(\lambda_m, T) = \{0\}.$$

Hence the sum is direct:

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

Conclusion: The vector space V decomposes as a direct sum of eigenspaces if and only if the sum of those eigenspaces equals V. □

The Minimal Polynomial and Eigenspaces

The **minimal polynomial** $p_T(z)$ is defined as the **monic polynomial of lowest degree** such that:

$$p_T(T) = 0.$$

Unlike the characteristic polynomial, which may include repeated roots corresponding to eigenvalues with algebraic multiplicity greater than one, the minimal polynomial of a diagonalizable operator contains only distinct linear factors. If and only if T is diagonalizable, the minimal polynomial splits into linear factors with no repeated roots.

$$p_T(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m).$$

Each root λ_i corresponds to an eigenvalue of T, and **no factor appears with an exponent greater than 1**. This ensures that $p_T(z)$ is the simplest polynomial that completely annihilates T.

If the minimal polynomial contains a repeated factor such as $(z-\lambda)^2$, this implies that T is not diagonalizable. Indeed, by the Cayley–Hamilton Theorem (Theorem 4.4.2), every operator satisfies its own characteristic polynomial. This means that the characteristic polynomial always annihilates T. However, even if the characteristic polynomial contains a repeated linear factor such as $(z-\lambda)^2$, this does not imply that every polynomial which annihilates T - including the minimal polynomial - must also contain that factor with the same multiplicity. The multiplicity of factors in the minimal polynomial may differ, and it is this structure that determines diagonalizability. Thus, one must examine the minimal polynomial directly - diagonalizability fails only when it contains repeated linear factors.

Multiplicity and Diagonalizability

As defined earlier in Definitions 2.1.10 and 2.1.11, each eigenvalue λ of T gives rise to two key notions of multiplicity:

- The *algebraic multiplicity* is the multiplicity of λ as a root of the characteristic polynomial.
- The *geometric multiplicity* is the dimension of the eigenspace $E(\lambda, T) = \ker(T - \lambda I)$.

An operator $T \in \mathcal{L}(V)$ is diagonalizable if and only if:

1. The characteristic polynomial splits into linear factors over the field.
2. For every eigenvalue λ, the geometric multiplicity equals its algebraic multiplicity:

$$\dim E(\lambda, T) = \text{mult}_{\text{alg}}(\lambda).$$

This ensures that there are enough linearly independent eigenvectors to form a basis of V, allowing T to be diagonalized.

Application: Invariant Subspaces and Stability Consider the restriction of T to an invariant subspace $U \subseteq V$. If U is associated with eigenvalues $\{\lambda_1, \lambda_2\}$, the restriction $T|_U$ remains diagonalizable, illustrating the stability of eigenspace decomposition under such constraints:

$$U = E(\lambda_1, T|_U) \oplus E(\lambda_2, T|_U).$$

This provides a concrete framework for analyzing substructures within V.

Limitations and Generalization: When eigenvalues are not distinct or when eigenspaces do not span V, the direct sum property fails. This necessitates the introduction of generalized eigenspaces and Jordan Normal Form, as explored in Chapter 4.

Summary. This section established when a linear operator can be diagonalized by showing how eigenspaces may form a direct sum of the entire space. Diagonalizability depends on both the structure of the minimal polynomial and the equality between geometric and algebraic multiplicities. When these conditions fail, the operator is not diagonalizable — setting the stage for the generalized framework of Jordan Normal Form.

3.5 Conditions for Diagonalizability

In the previous section, we saw that if a linear operator has enough linearly independent eigenvectors to form a basis for the vector space, then the operator is diagonalizable. This happens precisely when the space decomposes as a direct sum of its eigenspaces.

However, not all operators satisfy these conditions. In this section, we examine the precise algebraic criteria that determine whether or not a linear operator is diagonalizable. In particular, we investigate the roles of the characteristic and minimal polynomials, and the relationship between algebraic and geometric multiplicity of eigenvalues.

The goal is to establish necessary and sufficient conditions for diagonalizability, and to understand what breaks down when those conditions fail — a theme that naturally leads into the study of generalized eigenspaces and Jordan decomposition in the next chapter.

Theorem 3.5.1 (Equivalent Conditions for Diagonalizability, cf. Axler (2024), p. 165)**.** *Let $T \in \mathcal{L}(V)$, where V is finite-dimensional, and let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. The following are equivalent:*

1. *T is diagonalizable.*
2. *V has a basis of eigenvectors of T.*
3. *$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.*
4. *$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.*

Proof. **(1) $\Rightarrow$ (2):** If T is diagonalizable, then by definition there exists a basis for V consisting entirely of eigenvectors of T.

(2) $\Rightarrow$ (3): A basis of eigenvectors means that V is the sum of the corresponding eigenspaces. Since the eigenvalues are distinct, the eigenspaces are linearly independent, and the sum is direct:

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

(3) $\Rightarrow$ (4): The dimension of a direct sum equals the sum of the dimensions of its components. Therefore,

$$\dim V = \sum_{i=1}^{m} \dim E(\lambda_i, T).$$

(4) $\Rightarrow$ (1): If the eigenspaces span V and their dimensions sum to $\dim V$, then their union is a linearly independent spanning set — that is, a basis for V. Thus, T is diagonalizable.

Conclusion: Each statement implies the next, forming a logical cycle. Hence, all four conditions are equivalent. □

Theorem 3.5.2 (Minimal Polynomial and Diagonalizability)**.** *$T \in \mathcal{L}(V)$ is diagonalizable if and only if its minimal polynomial factors as:*

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m),$$

where $\lambda_1, \ldots, \lambda_m$ *are distinct eigenvalues.*

Proof. **Sufficient Condition (⇒):** Assume T is diagonalizable. Then V has a basis of eigenvectors of T, and T can be represented as a diagonal matrix D in this basis:

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T. The minimal polynomial $p(z)$ is the unique monic polynomial of smallest degree such that $p(T) = 0$. Since T is diagonal, $p(T) = 0$ implies:

$$p(z) = (z - \lambda_1)^{m_1} \cdots (z - \lambda_k)^{m_k},$$

where $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues, and $m_i = 1$ because there are no generalized eigenvectors for T (as T is diagonalizable). Therefore:

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m),$$

with $\lambda_1, \ldots, \lambda_m$ distinct.

Necessary Condition (⇐): Suppose the minimal polynomial of T factors as:

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m),$$

where $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues. By the Cayley-Hamilton theorem 4.4.2, T satisfies its characteristic polynomial, which is a multiple of the minimal polynomial:

$$c(z) = (z - \lambda_1)^{k_1} \cdots (z - \lambda_m)^{k_m},$$

where $k_i \geq 1$. The distinct factors of $p(z)$ imply that the eigenspaces $E(\lambda_i, T) = \ker(T - \lambda_i I)$ span V and are pairwise disjoint:

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

Since the eigenspaces cover V, T is diagonalizable because there exists a basis of eigenvectors for T, each corresponding to an eigenvalue λ_i.

Conclusion: The minimal polynomial $p(z)$ factors into distinct linear terms if and only if T is diagonalizable, completing the proof. □

Expanded Discussion: Minimal Polynomial and Multiplicities When eigenvalues have higher algebraic multiplicities, diagonalizability depends on whether the geometric multiplicity of each eigenvalue equals its algebraic multiplicity. For example:

$$p_T(z) = (z - \lambda_1)^2(z - \lambda_2),$$

requires $\dim E(\lambda_1, T) = 2$ and $\dim E(\lambda_2, T) = 1$ for T to be diagonalizable.

However, **the presence of repeated eigenvalues does not necessarily imply non-diagonalizability**. A matrix with repeated eigenvalues is still diagonalizable if and only if each eigenvalue's **geometric multiplicity matches its algebraic multiplicity**:

$$\dim E(\lambda_i, T) = k_i, \quad \text{where } k_i \text{ is the algebraic multiplicity of } \lambda_i.$$

For instance, the matrix

$$M = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

is **diagonalizable** even though $\lambda = 3$ appears twice, because it has two linearly independent eigenvectors.

Conversely, the matrix

$$J = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

is **not diagonalizable**, since it has only one linearly independent eigenvector despite $\lambda = 3$ having algebraic multiplicity 2. This necessitates a Jordan block representation instead.

Thus, **the key criterion for diagonalization is the match between geometric and algebraic multiplicity, not merely the repetition of eigenvalues**.

Theorem 3.5.3 (Restriction to Invariant Subspaces, cf. Axler (2024, Thm. 5.65, p. 170))**.** *Let $T \in \mathcal{L}(V)$ be diagonalizable, and let $U \subseteq V$ be a T-invariant subspace. Then the restriction $T|_U$ is diagonalizable.*

Proof. Since T is diagonalizable, there exists a basis $\{v_1, \ldots, v_n\}$ of V consisting of eigenvectors of T, so $T(v_i) = \lambda_i v_i$ for each i.

Because U is T-invariant, we have $T(v_i) \in U$ for each $v_i \in U$. Hence, if we select those basis vectors v_i that lie in U, they are eigenvectors of $T|_U$, and they span U.

Therefore, U admits a basis of eigenvectors for $T|_U$, and so $T|_U$ is diagonalizable. $\square$

Summary. A linear operator is diagonalizable if its characteristic polynomial splits into linear factors and the geometric multiplicity of each eigenvalue equals its algebraic multiplicity. The minimal polynomial must then have no repeated factors. When these conditions fail, diagonalization is no longer possible — and Jordan Normal Form becomes the natural generalization.

Transition to Jordan Normal Form. The results of this chapter show that diagonalization hinges on the existence of enough linearly independent eigenvectors. However, when this condition is not met — when, for example, the geometric multiplicity of an eigenvalue is strictly less than its algebraic multiplicity diagonalization fails. In such cases, we require a broader framework. The Jordan Normal Form provides precisely this: a canonical structure that generalizes diagonalization by introducing chains of generalized eigenvectors and nilpotent components. What diagonalization cannot capture, the Jordan form reveals.

4 Jordan Normal Form

4.1 Overview

In Chapter 3, we developed the theory of diagonalization and showed how linear operators can be simplified via a change of basis—provided that there are sufficiently many linearly independent eigenvectors. However, not all linear operators admit such a simplification. In some cases, the geometric multiplicity of an eigenvalue is strictly less than its algebraic multiplicity, which implies that the matrix cannot be diagonalized.

For example, consider the matrix

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

The only eigenvalue is $\lambda = 3$, and the corresponding eigenspace is one-dimensional, despite its algebraic multiplicity being two. Hence, A is not diagonalizable. Still, A exhibits highly structured behavior—it can be transformed into a simpler form that captures its essential algebraic and geometric features.

To systematically handle such non-diagonalizable operators, we require a more general canonical representation: the *Jordan Normal Form.* This representation generalizes diagonalization by incorporating both eigenvectors and so-called generalized eigenvectors, organized into structures known as *Jordan blocks.* The resulting matrix is nearly diagonal—block upper triangular with eigenvalues on the diagonal and ones on the superdiagonal—and uniquely determined up to similarity.

The goal of this chapter is to construct Jordan Normal Form and explain its theoretical foundation. We proceed in several stages:

- We begin with the notion of generalized eigenvectors and the role of nilpotent operators in shaping the internal algebraic structure of linear operators.
- We then establish the decomposition of the vector space into generalized eigenspaces, which provides the scaffolding for assembling Jordan blocks.
- The chapter culminates with the formal definition of Jordan Normal Form, including concrete examples and structural consequences.

Summary. This chapter introduces the core motivation behind Jordan Normal Form. While diagonalization works for operators with a complete basis of eigenvectors, many linear transformations are not diagonalizable. Jordan Normal Form provides a canonical structure that extends diagonalizability to the broader class of linear operators over $\mathbb{C}$, capturing both eigenvalues and nilpotent behavior.

4.2 Generalized Eigenvectors and Nilpotent Operators

Introduction

To construct the Jordan Normal Form, we rely on the concepts of **generalized eigenvectors** and **nilpotent operators**, which provide the underlying algebraic structure necessary to analyze non-diagonalizable matrices.

Standard eigenvectors satisfy: $(T - \lambda I)v = 0$. However, when this equation does not produce enough independent vectors, we must refine our approach. Generalized eigenvectors allow higher powers of $(T - \lambda I)$ to act before vanishing, effectively **bridging the gap** between diagonalizable and non-diagonalizable cases.

Null Spaces of Powers of an Operator

Let T be a linear operator on a vector space V. The **null spaces of increasing powers** of T, defined as:

$$\text{null}(T^k) = \{v \in V \mid T^k v = 0\}.$$

We recognize that, in the context of Jordan block construction, the operator T is typically considered as $T - \lambda I$, for a fixed eigenvalue λ. Thus, the null spaces above reflect those of increasing powers of the nilpotent operator $T - \lambda I$.

This operator T can now form an ascending sequence:

$$\{0\} \subseteq \text{null}(T) \subseteq \text{null}(T^2) \subseteq \cdots \subseteq \text{null}(T^k).$$

By the Fundamental Theorem of Linear Maps (Theorem 3.2.1), this sequence stabilizes for some integer m: $\text{null}(T^m) = \text{null}(T^{m+1}) = \ldots$. This stabilization is fundamental, as it ensures that a finite-dimensional space has a well-defined limit, allowing us to systematically determine the generalized eigenvector chains that form Jordan blocks. Vectors that seem new to null(T^{m+1}) are already in null(T^m) - stabilization means that no additional contributions will be added.

Generalized Eigenvectors

Definition 4.2.1 (Generalized Eigenvector, cf. Axler (2024, Def. 8.8, p. 300))**.** *Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a generalized eigenvector of T corresponding to λ if $v \neq 0$ and $(T - \lambda I)^k v = 0$ for some positive integer k.*

Unlike regular eigenvectors, **generalized eigenvectors** allow higher powers of $(T - \lambda I)$ to act before vanishing, capturing additional algebraic structure of T that enables the formation of Jordan blocks.

Chains of Generalized Eigenvectors. A generalized eigenvector v_k corresponding to an eigenvalue λ satisfies

$$(T - \lambda I)^k v_k = 0 \quad \text{but} \quad (T - \lambda I)^{k-1} v_k = v_1 \neq 0,$$

where v_1 serves as the associated (standard) eigenvector.

This means that applying $(T - \lambda I)$ repeatedly yields a chain of vectors:

$$v_k \xmapsto{(T-\lambda I)} v_{k-1} \xmapsto{(T-\lambda I)} \cdots \xmapsto{(T-\lambda I)} v_2 \xmapsto{(T-\lambda I)} v_1 \xmapsto{(T-\lambda I)} 0,$$

where each $v_i = (T - \lambda I)^{k-i} v_k$. These vectors form a *generalized eigenvector chain*, and will later appear as columns in a Jordan block associated with λ.

Nilpotent Operators

Definition 4.2.2. *An operator T is said to be* ***nilpotent*** *if there exists $m \in \mathbb{N}$ such that:*

$$T^m = 0.$$

Having defined nilpotent operators, we next examine how rapidly such operators "collapse" vectors to zero. Since applying T repeatedly must eventually annihilate every vector, it is natural to ask how many steps are needed at most. The following theorem shows that this number is always bounded by $\dim V$.

Theorem 4.2.3 (Nilpotent Operator Theorem, cf. Axler (2024, Thm. 8.16, p. 304))**.** *Let* $T \in \mathcal{L}(V)$ *be nilpotent and* $\dim V = n$*. Then*

$$T^n = 0.$$

Proof. Because T is nilpotent, there exists a positive integer k such that $T^k = 0$, and hence $\operatorname{null}(T^k) = V$. As noted in this section, the sequence $\operatorname{null}(T^j)$ stabilizes at or before $j = \dim V$. Thus, $\operatorname{null}(T^{\dim V}) = V$, implying $T^{\dim V} = 0$. □

Key Properties of Nilpotent Operators

- The minimal polynomial of a nilpotent operator T has the form x^m.
- The only eigenvalue of a nilpotent operator is 0.

Since the **nilpotent part** of $(T - \lambda I)$ determines Jordan block structure, understanding nilpotency is **crucial** in our decomposition.

Example 4.2.4 (Generalized Eigenvectors and Nilpotent Operators)**.** *Consider the matrix:*

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

- *Eigenvalue:* $\lambda = 2$.
- *The eigenspace is* ***one-dimensional****, meaning diagonalization is impossible.*
- *A second vector satisfies* $(A - 2I)^2 v = 0$*, confirming its* ***generalized eigenvector*** *nature.*
- *The Jordan block for this matrix has size* 3*, linking eigenvectors into a sequence via nilpotency.*

This example illustrates how generalization is ***essential*** *when diagonalization fails.*

This space contains all eigenvectors and generalized eigenvectors corresponding to λ, and forms the structural foundation for decomposing V into invariant subspaces. We now turn to this decomposition.

Summary. This section introduces generalized eigenvectors and nilpotent operators as tools to describe the structure of non-diagonalizable transformations. Generalized eigenvectors form chains under $(T - \lambda I)$, and nilpotent operators exhibit this chain structure algebraically. These concepts lay the groundwork for the decomposition of V into generalized eigenspaces.

4.3 Generalized Eigenspace Decomposition

Introduction

In the previous section, we introduced generalized eigenvectors and saw how they arise naturally in the study of non-diagonalizable operators. These vectors form chains under the action of nilpotent parts of T, capturing richer structure than standard eigenvectors alone.

By organizing these vectors into subspaces—called *generalized eigenspaces*—we gain access to a decomposition of V into invariant subspaces aligned with each eigenvalue. This structural insight reveals the internal block-diagonal form of T, as realized in its Jordan Normal Form.

Definition 4.3.1 (Generalized Eigenspace, cf. Axler (2024, Def. 8.19, p. 308))**.** *Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The* generalized eigenspace *of T corresponding to λ is defined as*

$$G(\lambda, T) = \left\{ v \in V \mid (T - \lambda I)^k v = 0 \text{ for some positive integer } k \right\}.$$

This subspace contains both eigenvectors and generalized eigenvectors corresponding to λ, and thus extends beyond the standard eigenspace. It will serve as a fundamental building block in decomposing V into T-invariant subspaces.

Theorem 4.3.2 (Generalized Eigenspace as Null Space, cf. Axler (2024, Thm. 8.20, p. 308))**.** *Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then*

$$G(\lambda, T) = \operatorname{null}\left((T - \lambda I)^{\dim V}\right).$$

Proof. Suppose $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. By the definition of generalized eigenspace, this implies $v \in G(\lambda, T)$, so

$$G(\lambda, T) \supseteq \operatorname{null}(T - \lambda I)^{\dim V}.$$

Conversely, suppose $v \in G(\lambda, T)$. Then there exists a positive integer k such that $(T-\lambda I)^k v = 0$, i.e., $v \in \operatorname{null}(T-\lambda I)^k$. But since the null spaces of powers of a linear operator form an ascending chain that stabilizes at $\dim V$ (see Section 4.2), it follows that

$$v \in \operatorname{null}(T-\lambda I)^{\dim V}.$$

Hence,

$$G(\lambda, T) \subseteq \operatorname{null}(T-\lambda I)^{\dim V},$$

which completes the proof. □

Remark. *This result has two important consequences. First, it provides a practical method for computing generalized eigenspaces: instead of checking multiple powers of $(T-\lambda I)$, it suffices to compute a single null space. Second, it guarantees that $G(\lambda, T)$ is a T-invariant subspace of V, since null spaces of linear operators are always invariant. These properties are fundamental in constructing the decomposition described in following theorem.*

Theorem 4.3.3 (Generalized Eigenspace Decomposition, cf. Axler (2024, Thm. 8.22, p. 309))**.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then:*

1. *Each generalized eigenspace $G(\lambda_k, T)$ is invariant under T.*
2. *The operator $(T-\lambda_k I)\big|_{G(\lambda_k, T)}$ is nilpotent.*
3. *The space V is the direct sum:*

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T).$$

Proof. **(1)** By Theorem 4.3.2, we have

$$G(\lambda_k, T) = \operatorname{null}\left((T-\lambda_k I)^{\dim V}\right).$$

Since null spaces of powers of linear operators are invariant under the operator itself, it follows that $G(\lambda_k, T)$ is invariant under T.

(2) Let $v \in G(\lambda_k, T)$. Then, by definition, there exists a positive integer p such that

$$(T-\lambda_k I)^p v = 0.$$

Thus, $(T-\lambda_k I)$ acts nilpotently on every vector in $G(\lambda_k, T)$. More precisely, by Theorem 4.3.2, we have

$$G(\lambda_k, T) = \text{null}(T - \lambda_k I)^{\dim V},$$

so the restriction

$$\left((T - \lambda_k I)\big|_{G(\lambda_k,T)}\right)^{\dim V} = 0,$$

which confirms that the restriction is nilpotent.

(3) Since $\mathbb{F} = \mathbb{C}$, the minimal polynomial of T splits into linear factors:

$$m_T(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_m)^{n_m},$$

where the $(x - \lambda_k)^{n_k}$ are pairwise relatively prime. By the primary decomposition theorem, V decomposes as a direct sum of the null spaces:

$$V = \text{null}(T - \lambda_1 I)^{n_1} \oplus \cdots \oplus \text{null}(T - \lambda_m I)^{n_m}.$$

From Theorem 4.3.2, we know that

$$G(\lambda_k, T) = \text{null}(T - \lambda_k I)^{\dim V},$$

and since $n_k \leq \dim V$, we have

$$\text{null}(T - \lambda_k I)^{n_k} \subseteq G(\lambda_k, T).$$

Moreover, the generalized eigenspaces are disjoint and together span V, so we conclude:

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T).$$

□

This decomposition provides the structural foundation for Jordan Normal Form. Since each subspace $G(\lambda_k, T)$ is invariant under T, and the restriction of T to each is nilpotent modulo scalar shift, the action of T becomes block-diagonal when expressed in a basis adapted to this decomposition. The size and number of Jordan blocks within each generalized eigenspace are then determined by the nilpotent action of $(T - \lambda_k I)$.

Multiplicity of Eigenvalues

The **algebraic multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic polynomial of T. In this setting, the algebraic multiplicity of λ coincides with $\dim G(\lambda, T)$, since the minimal polynomial splits into linear factors and V decomposes as a direct sum of generalized eigenspaces.

The sum of these multiplicities across all distinct eigenvalues satisfies:

$$\sum_{i=1}^{m} \dim G(\lambda_i, T) = \dim V.$$

This guarantees that the decomposition from Theorem 4.3.3 is exhaustive: every vector in V belongs to one of the generalized eigenspaces.

Block-Diagonal Representation and Jordan Blocks

Since V decomposes into T-invariant generalized eigenspaces, the matrix representation of T in a suitable basis becomes block-diagonal:

$$J = J_{\lambda_1} \oplus J_{\lambda_2} \oplus \cdots \oplus J_{\lambda_m},$$

where each block J_{λ_i} corresponds to a generalized eigenspace $G(\lambda_i, T)$.

Within each block, the structure is governed by the nilpotency of $(T - \lambda I)$ and consists of one or more **Jordan blocks**, each of the form:

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}.$$

A Jordan block of size k reflects a chain of generalized eigenvectors

$$v_k \mapsto v_{k-1} \mapsto \cdots \mapsto v_1 \mapsto 0,$$

where each arrow corresponds to the action of $(T - \lambda I)$. This structure encodes the nilpotency of $T - \lambda I$ on $G(\lambda, T)$ and gives rise to the superdiagonal ones in the matrix.

Example: Generalized Eigenspace Decomposition in Action

Example 4.3.4. *Consider the matrix*

$$A = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Step 1: Compute Eigenvalues. *The characteristic polynomial is*

$$\det(A - \lambda I) = (4 - \lambda)^2 (3 - \lambda)^2.$$

Thus, the eigenvalues are $\lambda_1 = 4$ *and* $\lambda_2 = 3$*, both with algebraic multiplicity 2.*

Step 2: Compute Generalized Eigenspaces. *For* $\lambda_1 = 4$*, we compute:*

$$G(4, A) = \mathrm{null}((A - 4I)^2).$$

For $\lambda_2 = 3$*, similarly:*

$$G(3, A) = \mathrm{null}((A - 3I)^2).$$

Each generalized eigenspace is two-dimensional, whereas the corresponding eigenspaces are one-dimensional. This indicates the existence of generalized eigenvector chains of length 2 *for each eigenvalue, and reveals that the algebraic multiplicity is* 2 *but the geometric multiplicity is* 1 *in each case.*

Step 3: Construct Jordan Normal Form. *The Jordan form of* A *is:*

$$J = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

We see that the Jordan form consists of:

- *One* 2×2 *Jordan block for eigenvalue 4*
- *One* 2×2 *Jordan block for eigenvalue 3*

This example illustrates how generalized eigenspace decomposition leads directly to Jordan blocks and reveals the internal structure of the operator.

Summary. The generalized eigenspace decomposition reveals the internal structure of any linear operator $T \in \mathcal{L}(V)$ over $\mathbb{C}$. By partitioning V into invariant subspaces $G(\lambda, T)$, where each restriction of T becomes nilpotent modulo scalar shift, we achieve a block-diagonal representation of T. Within each block, Jordan substructures emerge naturally through chains of generalized eigenvectors.

This provides not only a refined classification of operators but also a practical framework for computation. The resulting Jordan Normal Form encapsulates both spectral information and the nilpotent behavior of T, forming a bridge between algebraic theory and concrete applications in dynamics, matrix functions, and beyond.

4.4 Consequences of Generalized Eigenspace Decomposition

Introduction

The generalized eigenspace decomposition provides a powerful framework for understanding linear operators. By expressing V as a direct sum of generalized eigenspaces, we gain insight into both the eigenvalue structure and the internal nilpotent behavior of T. This decomposition not only underlies the construction of Jordan Normal Form but also enables precise applications, such as computing functions of operators.

This section presents three major consequences of the generalized eigenspace decomposition. First, we prove that every linear operator satisfies its own characteristic polynomial — the celebrated Cayley–Hamilton Theorem, which connects algebraic and spectral structure.

Second, we show that every invertible operator admits a square root under certain structural conditions, providing insight into the functional calculus of operators.

Third, we examine nilpotent operators and their finite iteration behavior, shedding light on the role of nilpotent blocks in Jordan matrices and the finite structure of generalized eigenvector chains.

These results collectively build the foundation for the operator calculus developed in the following sections.

Cayley–Hamilton Theorem

Example 4.4.1 (Motivating Cayley–Hamilton's Theorem)**.** *Consider the matrix* $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. *Its characteristic polynomial is* $p_A(x) = (x-2)^2$, *and one can verify that* $p_A(A) = 0$. *This illustrates that matrices may satisfy their own characteristic polynomial even when not diagonalizable — a phenomenon formalized in following theorem.*

Theorem 4.4.2 (Cayley–Hamilton, cf. Axler (2024, Thm. 8.29, p. 312))**.** *Let* $T \in \mathcal{L}(V)$ *be a linear operator on a finite-dimensional vector space. Then*

$$p_T(T) = 0,$$

where $p_T(x) = \det(xI - T)$ *is the characteristic polynomial of* T.

Proof (cf. Axler (2024), p. 312). Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, and let $d_k = \dim G(\lambda_k, T)$. For each $k \in \{1, \ldots, m\}$, we know from theorem 4.2.3 that

$$(T - \lambda_k I)^{d_k} \big[G(\lambda_k, T)\big] = 0.$$

The generalized eigenspace decomposition (Theorem 4.3.3) states that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T).$$

Hence, to show that $p_T(T) = 0$, it suffices to show that $p_T(T)[G(\lambda_k, T)] = 0$ for each k.

Recall that the characteristic polynomial is given by

$$p_T(x) = \prod_{k=1}^{m} (x - \lambda_k)^{d_k}.$$

Because the operators $T - \lambda_k I$ commute with one another, we can move the factor $(T - \lambda_k I)^{d_k}$ to the right in the product expression for $p_T(T)$. Then since

$$(T - \lambda_k I)^{d_k} [G(\lambda_k, T)] = 0,$$

it follows that $p_T(T)[G(\lambda_k, T)] = 0$ for each k. As V is the direct sum of the generalized eigenspaces, we conclude that $p_T(T) = 0$ on all of V. $\square$

Having established the foundational consequences of the generalized eigenspace decomposition, including the nilpotency of certain operator components and the Cayley-Hamilton Theorem, we are now prepared to explore how these structural insights enable concrete computations. In particular, we will show that every invertible operator in a complex vector space admits a square root. This result depends critically on the decomposition of V into generalized eigenspaces within which the functional calculus becomes tractable.

Remark. *Since the operator satisfies its characteristic polynomial according to the Cayley–Hamilton theorem, it also follows that it satisfies its minimal polynomial — which is a factor of the characteristic polynomial. This result provides structural information that is essential when analyzing candidates for the minimal polynomial.*

Remark. *In light of the Cayley-Hamilton Theorem, which asserts that every square matrix annihilates its own characteristic polynomial, we gain an essential structural insight: the product $(A - \lambda_1 I) \cdots (A - \lambda_n I)$, for distinct nonzero eigenvalues of an invertible matrix A, removes from the space all vectors associated with its eigendirections. That is, each factor $(A - \lambda_i I)$ corresponds to a "removal" of the eigenspace linked to λ_i, resulting in a cumulative filtering of all eigenstructure. While each individual $A - \lambda_i I$ is not the zero matrix, the total product becomes so — precisely because all directional components have been erased.*

This interpretation not only affirms the annihilating power of the minimal polynomial, but also anticipates our use of generalized eigenspace decomposition in the functional calculus. Specifically, the realization that V is built from eigenvector directions — each of which can be systematically accessed and "isolated" through such factorizations — underpins the tractability of functions of operators such as the square root.

Square Roots of Operators

We now present a theorem showing that every invertible operator on a complex vector space has a square root. This is a direct consequence of the generalized eigenspace decomposition and mirrors Theorem 8.39 in Axler (2024).

Theorem 4.4.3 (Square Root of Invertible Operator, cf. Axler (2024, Thm. 8.39, p. 319))**.** *Suppose $T \in \mathcal{L}(V)$ is invertible and $\mathbb{F} = \mathbb{C}$. Then there exists an operator $S \in \mathcal{L}(V)$ such that*

$$S^2 = T.$$

Proof. By the generalized eigenspace decomposition (Theorem 4.3.3), we can write:

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $\lambda_i \in \mathbb{C}$ is an eigenvalue of T and each $G(\lambda_i, T)$ is invariant under T. Since T is invertible, each eigenvalue $\lambda_i \neq 0$.

Restrict T to each generalized eigenspace $G(\lambda_i, T)$. Then T acts on this subspace as:

$$T_i = \lambda_i I + N_i,$$

where N_i is nilpotent. Since $\lambda_i \neq 0$, the operator $\lambda_i^{-1} T_i = I + \lambda_i^{-1} N_i$ is of the form $I + N$ with N nilpotent. For such operators, we can define a square root via a finite Taylor series:

$$(I + N)^{1/2} = \sum_{k=0}^{d} \binom{1/2}{k} N^k,$$

where d is the index of nilpotency of N. Applying this to $\lambda_i^{-1} T_i$ gives a square root of the restriction. Multiplying by $\sqrt{\lambda_i}$, we obtain an operator S_i such that $S_i^2 = T_i$.

Define $S = S_1 \oplus \cdots \oplus S_m$. Then $S^2 = T$, completing the proof. $\square$

Remark. *The proof shows that square roots of invertible operators can be constructed piecewise over each generalized eigenspace, reducing the problem to the analysis of nilpotent perturbations of identity. This highlights the utility of the decomposition not just for classification, but also for operator calculus.*

This theorem highlights how decomposition simplifies the analysis of matrix functions, particularly square roots.

Existence and Structure of Jordan Form

Definition 4.4.4 (Jordan Basis, cf. Axler (2024, Def. 8.44, p. 322))**.** *Let $T \in \mathcal{L}(V)$. A basis of V is called a* Jordan basis *for T if the matrix of T with respect to this basis is block-diagonal:*

$$\begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_p \end{pmatrix},$$

where each block J_k has the form

$$J_k = \begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \lambda_k & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k & 1 \\ 0 & 0 & \cdots & 0 & \lambda_k \end{pmatrix}.$$

Theorem 4.4.5 (Jordan Basis for Nilpotent Operator, cf. Axler (2024, Thm. 8.45, p. 322))**.** *Suppose $T \in \mathcal{L}(V)$ is nilpotent. Then there exists a basis of V that is a Jordan basis for T.*

Remark. *In this setting, each Jordan block has eigenvalue* 0 *and takes the form:*

$$J_k(0) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

This theorem establishes that even in the extreme case where an operator has only 0 as a root of its minimal polynomial — that is, when it is entirely nilpotent — the Jordan form still exists and consists of Jordan blocks with eigenvalue 0. The chains of generalized eigenvectors are simply constructed from the kernel hierarchy of T^k.

Proof. By definition of nilpotence, there exists a positive integer m such that $T^m = 0$. Consider the sequence of null spaces

$$\{0\} \subseteq \operatorname{null}(T) \subseteq \operatorname{null}(T^2) \subseteq \cdots \subseteq \operatorname{null}(T^m) = V,$$

which stabilizes at V since $T^m = 0$. For each $k = 1, \ldots, m$, define

$$W_k := \operatorname{null}(T^k) \setminus \operatorname{null}(T^{k-1}).$$

Each vector $v \in W_k$ satisfies $T^k v = 0$ but $T^{k-1} v \neq 0$, and thus gives rise to a chain

$$v, Tv, T^2 v, \ldots, T^{k-1} v,$$

consisting of k linearly independent generalized eigenvectors (since T acts nilpotently). These chains correspond to Jordan blocks with eigenvalue 0.

By choosing a basis for each W_k and forming chains as above, we obtain a basis of V consisting of generalized eigenvectors organized into Jordan chains. Thus, the matrix of T with respect to this basis is block-diagonal with Jordan blocks — all corresponding to the eigenvalue 0.

$\square$

Remark (Revisiting Chains and Blocks)**.** *The structural behavior of the operator seen in Example 4.3.4 fully illustrates the consequences of generalized eigenspace decomposition.*

The matrix there exhibits two eigenvalues with algebraic multiplicity 2 and geometric multiplicity 1, resulting in chains of generalized eigenvectors and the emergence of 2×2 *Jordan blocks.*

Since this example already demonstrates how limited geometric multiplicity leads to block formation, we omit an additional repetition here.

Instead, we note:

- *The structure of V is partitioned into generalized eigenspaces for each eigenvalue.*
- *Each eigenspace gives rise to a block reflecting internal nilpotency.*
- *Jordan chains emerge naturally from the need to extend eigenvectors into complete bases.*

For details, see Example 4.3.4, which confirms all steps leading to Jordan Normal Form.

This example shows how generalized eigenspaces directly determine Jordan blocks.

Theorem 4.4.6 (Existence of Jordan Basis, cf. Axler (2024, Thm. 8.46, p. 324))**.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then there exists a basis of V that is a Jordan basis for T.*

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. The generalized eigenspace decomposition (Theorem 4.3.3) gives

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each restriction $(T - \lambda_k I)\big|_{G(\lambda_k,T)}$ is nilpotent. By theorem 4.4.5, there exists a basis of each $G(\lambda_k, T)$ that is a Jordan basis for this restriction. Combining these bases yields a Jordan basis for T. $\square$

Summary. The generalized eigenspace decomposition enables a deep structural understanding of linear operators. It not only leads to Jordan Normal Form — a canonical representation that captures both eigenvalues and nilpotent behavior — but also facilitates functional analysis of operators. In particular, square roots of invertible operators can be constructed by working within each generalized eigenspace. Moreover, every operator satisfies its characteristic polynomial (Cayley–Hamilton), and every nilpotent operator vanishes after finitely many iterations — both results made transparent by the decomposition. Together, these insights show that all operators over $\mathbb{C}$ admit a Jordan form built from local nilpotent structures.

With a structured understanding of Jordan decomposition, we now turn to a fundamental invariant: the trace function. By examining how trace interacts with Jordan blocks, we can further reveal key algebraic properties of non-diagonalizable matrices.

4.5 Trace: A Connection Between Matrices and Operators

Introduction

The trace of a linear operator is a fundamental algebraic invariant that reflects key structural properties of matrix transformations. It remains unchanged under similarity transformations, making it a natural tool for studying canonical forms — especially Jordan matrices. Trace not only encodes the total eigenvalue sum but also resists distortion from nilpotent perturbations, thereby retaining spectral information even in non-diagonalizable settings.

In this section, we explore:

- The definition and fundamental properties of trace
- Its direct connection to eigenvalues
- Its role in interpreting Jordan block structure

Definition and Fundamental Properties

Trace is a fundamental operation that distills spectral information from matrices and linear operators. It is defined in purely algebraic terms, yet its implications reach deep into spectral theory. One of its most powerful consequences is the direct link between trace and the sum of eigenvalues.

Definition 4.5.1. *Let A be an $n \times n$ matrix over a field $\mathbb{F}$. The* trace *of A is defined as*

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

where a_{ii} denotes the entries along the main diagonal.

For a linear operator $T \in \mathcal{L}(V)$, the trace is independent of basis:

$$\mathrm{tr}(T) := \mathrm{tr}(M(T)),$$

where $M(T)$ is any matrix representation of T. That is, trace is an intrinsic property of the operator itself.

Key Properties of Trace

- **Invariance under Basis Change:** If A and B are similar, then $\mathrm{tr}(A) = \mathrm{tr}(B)$.
- **Linearity:** For any operators S, T and scalar α,

$$\mathrm{tr}(S + T) = \mathrm{tr}(S) + \mathrm{tr}(T), \qquad \mathrm{tr}(\alpha T) = \alpha\,\mathrm{tr}(T).$$

- **Cyclic Property:** For matrices A, B of compatible size,

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Trace as the Sum of Eigenvalues

The trace not only describes a structural property of a matrix — it also encodes precise spectral information. In particular, the trace of a linear operator equals the sum of its eigenvalues, counted with multiplicities.

Theorem 4.5.2 (cf. Axler (2024, Section 8D, p. 326))**.** *Let $T \in \mathcal{L}(V)$ be a linear operator with eigenvalues $\lambda_1, \ldots, \lambda_n$ (counted with multiplicities). Then:*

$$\mathrm{tr}(T) = \sum_{i=1}^{n} \lambda_i.$$

Remark. *Although this result is not numbered as a formal theorem in Axler (2024), it appears in Section 8D (p. 326) as a key observation: "The trace of a linear operator equals the sum of its eigenvalues, counted with multiplicities." In this work, we restate it as a theorem to highlight its central importance in spectral analysis.*

Proof. Let $M(T)$ denote a matrix representation of T in any basis. Since trace is invariant under basis change, it suffices to prove the statement for $M(T)$.

The characteristic polynomial is:

$$p_T(x) = \det(xI - M(T)) = x^n - (\lambda_1 + \cdots + \lambda_n)x^{n-1} + \cdots ,$$

where the coefficient of x^{n-1} equals $-\sum \lambda_i$ by the fundamental theorem of symmetric polynomials. Therefore:

$$\operatorname{tr}(M(T)) = \sum_{i=1}^{n} \lambda_i = \operatorname{tr}(T).$$

□

This result highlights that the trace gives immediate access to the sum of eigenvalues — without the need to compute individual eigenvectors or perform full diagonalization.

Trace in Jordan Normal Form

The trace of a Jordan matrix is particularly significant because Jordan blocks preserve eigenvalues while introducing nilpotency. This structure ensures that trace remains a direct measure of the total eigenvalue weight — even in the absence of diagonalizability.

Theorem 4.5.3 (Trace and Non-Diagonalizability, cf. Axler (2024, Sec. 8D))**.** *Let $T \in \mathcal{L}(V)$ be a linear operator over $\mathbb{C}$, with eigenvalues $\lambda_1, \ldots, \lambda_n$ counted with multiplicities. Then:*

$$\operatorname{tr}(T) = \sum_{i=1}^{n} \lambda_i,$$

regardless of whether T is diagonalizable or not.

Remark. *Although this result is not stated as a formal theorem in Axler (2024), it appears repeatedly as a key principle in Section 8D (see p. 327 and Sats 8.52). There, Axler emphasizes that the trace of an operator equals the* *sum of its eigenvalues — even in cases where the matrix is not diagonalizable. This is possible because Jordan matrices retain all eigenvalues on their diagonals, and trace depends only on these diagonal entries. The nilpotent structure (superdiagonal 1's) contributes nothing to the trace.*

Theorem 4.5.4 (Trace of Jordan Matrix, cf. Axler (2024, Sec. 8D))**.** *Let J be a Jordan matrix consisting of blocks J_{λ_i}. Then:*

$$\operatorname{tr}(J) = \sum_i k_i \lambda_i,$$

where k_i is the size of the Jordan block associated with eigenvalue λ_i.

Proof. Each Jordan block $J_k(\lambda)$ has λ along the diagonal and ones on the superdiagonal. Thus, its trace is:

$$\operatorname{tr}(J_k(\lambda)) = k\lambda.$$

Summing across all Jordan blocks yields:

$$\operatorname{tr}(J) = \sum_i k_i \lambda_i.$$

□

This result confirms that trace remains invariant under similarity transformations — and that it measures total eigenvalue weight, regardless of whether the operator is diagonalizable.

Example 4.5.5 (Computing Trace in Jordan Form)**.** *Consider the matrix:*

$$A = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Step 1: Compute the Trace *Trace is the sum of diagonal entries:*

$$\operatorname{tr}(A) = 4 + 4 + 3 + 3 = 14.$$

Step 2: Verify Using Eigenvalues *Eigenvalues are $\lambda_1 = 4$ (multiplicity 2) and $\lambda_2 = 3$ (multiplicity 2), so:*

$$\operatorname{tr}(A) = 4 \times 2 + 3 \times 2 = 14.$$

Both methods yield the same result — confirming that trace accounts for eigenvalue multiplicity even when the matrix is not diagonalizable.

Example 4.5.6. *The same applies to matrices that are not upper triangular.*
$A = PJP^{-1}, \begin{pmatrix} 5 & -1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}^{-1} \Rightarrow$

$$\operatorname{Tr}(A) = 5 + 1 = 3 + 3 = 2 \times 3 = \operatorname{Tr}(J)$$

Summary. The trace of a linear operator captures essential spectral information in an elegant and basis-independent way. It equals the sum of the operator's eigenvalues, counted with multiplicities, even when the operator is not diagonalizable. Jordan matrices retain this property: their nilpotent structure contributes nothing to the trace, as all eigenvalue information resides on the diagonal. This makes trace a robust invariant — equally valid in both diagonal and Jordan form — and a powerful analytic tool for studying linear transformations.

This section has shown that trace captures spectral information even in the presence of nilpotent structure. We now turn to the finer internal geometry of Jordan blocks — in particular, the reason why ones appear on their superdiagonal. Understanding this pattern reveals the deeper algebraic logic behind the form of non-diagonalizable operators.

4.6 Why this superdiagonal with ones?

Introduction

Although previous sections established the existence and properties of Jordan blocks, this section seeks to justify their internal structure from first principles. In particular, we analyze the mathematical necessity of the superdiagonal entries — showing how they emerge directly from invariance and non-diagonalizability.

To do so, we explore:
- The recursive relationship among generalized eigenvectors.
- Why an alternative transformation $Av_3 = \lambda v_3 + v_1$ fails to preserve the

required structure.
- How the correct form $Av_k = \lambda v_k + v_{k-1}$ naturally produces ones on the superdiagonal.

Motivating the Structure of Jordan Blocks Inductively

Step 1: Fundamental Properties of Invariant Subspaces

We begin by recognizing that a generalized eigenspace is an invariant subspace under A. This means that if A is not diagonalizable, then any change in the direction of a vector v under Av must occur through a vector that already belongs to the space.

The simplest case to analyze is a 2×2 matrix. Let A have an eigenvalue λ and a standard eigenvector v_1:

$$Av_1 = \lambda v_1.$$

However, since A is not diagonalizable, there must exist a linearly independent vector v_2 that does not satisfy a pure eigenvector relation. Instead, we reason as follows:

- Since the generalized eigenspace is invariant, A must shift v_2 while remaining within the space.
- The only available vector that can account for this shift is v_1.
- Therefore, the only valid transformation for v_2 must be:

$$Av_2 = \lambda v_2 + v_1.$$

Step 2: Why $Av_3 = \lambda v_3 + v_1$ Fails

We imagine that

$$Av_3 = \lambda v_3 + v_1.$$

If this were the case, it would mean that A acts on v_3 in exactly the same way as it acts on v_2, since we have already established that

$$Av_2 = \lambda v_2 + v_1.$$

But then the image of v_3 under A would be indistinguishable from that of v_2, and the only case in which this could happen is if $v_3 = v_2$. That contradicts the requirement that generalized eigenvectors in a Jordan chain must be linearly independent and therefore distinct. In particular, to construct a proper three-dimensional Jordan block, we must have three linearly independent vectors: v_1, v_2, v_3, each playing a unique role.

Thus, if v_3 is to belong to the same invariant subspace but remain distinct from v_2, A must act on it in a unique way — and the only possible change in direction is through v_2. Hence,

$$Av_3 = \lambda v_3 + v_2.$$

Step 3: Inductive Extension to v_k

Building on the established pattern, we conclude:

$$\begin{aligned} Av_2 &= \lambda v_2 + v_1, \\ Av_3 &= \lambda v_3 + v_2. \end{aligned}$$

By the same reasoning, each vector v_k for $k > 1$ must satisfy:

$$Av_k = \lambda v_k + v_{k-1}.$$

This recurrence guarantees that every vector in the Jordan chain is uniquely linked to its immediate predecessor, preserving the hierarchy of generalized eigenvectors and enabling the structured form of the Jordan block.

Step 4: Why This Structure Produces Superdiagonal Ones

The recurrence

$$Av_k = \lambda v_k + v_{k-1}$$

shows that each generalized eigenvector is directly linked to its immediate predecessor in the chain. In a matrix representation with respect to the ordered basis $\{v_1, v_2, \ldots, v_k\}$, this translates into a very specific structure:

- The diagonal entries must be λ, reflecting the eigenvalue associated with each v_i.
- The entry immediately above the diagonal (the superdiagonal) must be 1, since v_k maps to v_{k-1} with coefficient 1.

Therefore, the appearance of ones along the superdiagonal is not arbitrary: it is the inevitable matrix manifestation of the recurrence relation that defines the Jordan chain. This structure ensures that each step in the chain encodes a unique directional shift, preserving the hierarchy of generalized eigenvectors.

Remark (Explicit Matrix Realization of the Chain Condition)**.** *We recall the essential transformation step within a Jordan chain:*

$$Av_k = \lambda v_k + v_{k-1} \Rightarrow Av_k - \lambda v_k = v_{k-1} \Rightarrow (A - \lambda I)v_k = v_{k-1}.$$

This recurrence relation can be viewed through the lens of matrix structure. Consider:

$$(A - \lambda I)v_k = Av_k - \lambda v_k = PJP^{-1}v_k - \lambda v_k = PJe_k - \lambda v_k,$$

since $P^{-1}v_k = e_k$. Thus, PJe_k picks out the k-th column of PJ.

Now observe:

$$PJe_k - \lambda v_k = [\lambda v_1 \quad \lambda v_2 + v_1 \quad \dots \quad \lambda v_k + v_{k-1} \quad \dots \quad \lambda v_n + v_{n-1}]e_k - \lambda v_k = .$$

$$= \lambda v_k + v_{k-1} - \lambda v_k = v_{k-1}.$$

The reason we get a second term v_{k-1} in the k;th column of PJ is precisely because we have 1;s in the superdiagonal of J. The column in P that has the ordinal number $k-1$ is met by the 1; in the k;th column and row $k-1$ in J. We can now briefly summarize this as $(A - \lambda I)v_k = v_{k-1}$.

Link Between Generalized Eigenvectors and Matrix Representation

Let us now interpret the recurrence

$$Av_k = \lambda v_k + v_{k-1}$$

in matrix form. Define the matrix

$$P := [v_1 \ v_2 \ \dots \ v_n],$$

whose columns are the generalized eigenvectors forming a single Jordan chain of length n. Then, by construction, the action of A on the basis vectors satisfies:

$$AP = PJ,$$

where $J \in \mathbb{C}^{n \times n}$ is the Jordan block

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

Multiplying out PJ makes the recurrence relations manifest:

$$PJ = [\lambda v_1,\ \lambda v_2 + v_1,\ \lambda v_3 + v_2,\ \ldots,\ \lambda v_n + v_{n-1}],$$

which matches exactly with the action of A on the Jordan chain. Hence,

$$Av_k = \lambda v_k + v_{k-1} \quad \Longleftrightarrow \quad AP = PJ.$$

This validates that the Jordan matrix J is the correct matrix representation of A with respect to the basis $\{v_1, \ldots, v_n\}$, and confirms that the superdiagonal ones in J arise precisely because of the recurrence structure among the generalized eigenvectors.

Summary. The superdiagonal ones in a Jordan block are not arbitrary decorations — they are a structural necessity. The recurrence $Av_k = \lambda v_k + v_{k-1}$ ensures that each generalized eigenvector uniquely transforms into its predecessor. Any alternative would violate invariance and linear independence. Thus, the Jordan block's form reflects deep algebraic constraints imposed by non-diagonalizability.

Having understood the necessity of the Jordan block's internal structure, we now turn to a concrete example that illustrates how to compute a full Jordan decomposition in practice.

4.7 Practical Example: Using Jordan Normal Form

Introduction

This section focuses on practical examples. Although the approach is not entirely new at this point, we examine how a matrix A can be factored into PJP^{-1}. We then study a case where the matrix J partly consists of Jordan blocks, and where both standardized and generalized eigenvectors are chosen to build the matrix P.

By assembling these components into PJP^{-1}, which reproduces the original matrix A, we gain a deeper understanding of the bidirectional role of matrix factorization — that both $A \Rightarrow PJP^{-1}$ and $A \Leftarrow PJP^{-1}$ are possible paths.

The section concludes with a reflection on the possibility of choosing different bases for the same matrix. Although these considerations are not unique to non-diagonalizable matrices, the section is motivated by the ambition to provide a practical complement to the otherwise mainly theoretical perspective.

Example 4.7.1. *To illustrate how Jordan Normal Form helps analyze matrices that are not diagonalizable, consider the matrix:*

$$A = \begin{pmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{pmatrix}.$$

Our goal is to determine the Jordan Normal Form J of A by computing its eigenvalues, examining the structure of its eigenspaces and generalized eigenspaces, and constructing the transformation matrix P. This process reveals how nontrivial eigenstructure is systematically encoded through Jordan blocks.

—

Step 1: Compute the Characteristic and Minimal Polynomials

We begin by finding the characteristic polynomial of A:

$$p_A(x) = \det(A - xI) = (x-1)(x-2)(x-4)^2,$$

which shows that the eigenvalues are:

$$\lambda_1 = 1, \quad \lambda_2 = 2, \quad \lambda_3 = 4,$$

with $\lambda = 4$ having algebraic multiplicity 2.

This repeated factor suggests that A might not be diagonalizable. To investigate further, we compute the minimal polynomial:

$$m_A(x) = (x-1)(x-2)(x-4)^2.$$

Since the minimal polynomial includes $(x-4)^2$, we conclude that the largest Jordan block associated with $\lambda = 4$ must be of size 2×2. In particular:

- The exponent of $(x-4)$ in $m_A(x)$ tells us the size of the largest Jordan block for $\lambda = 4$.
*- The fact that this exponent is > 1 implies that A is **not diagonalizable**.*
- Therefore, $\lambda = 4$ must correspond to a Jordan block of size 2×2.

*We will soon confirm that the geometric multiplicity of $\lambda = 4$ is only 1, which confirms that we have **exactly one** such block.*

Step 2: Determine Eigenspaces and Generalized Eigenspaces

We now compute the (generalized) eigenspaces by solving $(A - \lambda I)v = 0$ for each eigenvalue.

- ***For** $\lambda = 1$**:** One eigenvector is*

$$v_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

- ***For** $\lambda = 2$**:** One eigenvector is*

$$v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

- ***For** $\lambda = 4$**:** Solving $(A - 4I)v = 0$, we find only a single eigenvector:*

$$v_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}.$$

The algebraic multiplicity of $\lambda = 4$ is 2, but we have found only one linearly independent eigenvector. Therefore, A is not diagonalizable and must include a generalized eigenvector associated with $\lambda = 4$.

To find it, we solve:

$$(A - 4I)^2 v_4 = 0 \quad \textit{with} \quad (A - 4I)v_4 = v_3.$$

This directly reflects the recurrence relation $Av_k = \lambda v_k + v_{k-1}$ discussed earlier.

A solution is:

$$v_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

Thus, the pair v_3, v_4 forms a Jordan chain:

$$(A - 4I)v_4 = v_3, \qquad (A - 4I)v_3 = 0,$$

which guarantees the presence of a Jordan block of size 2×2 for eigenvalue $\lambda = 4$.

Step 3: Construct the Jordan Normal Form

From Step 2, we determined that the eigenvalue $\lambda = 4$ has one eigenvector v_3 and one generalized eigenvector v_4, forming a Jordan chain of length 2. This implies that the Jordan Normal Form J must include a single Jordan block of size 2×2 for $\lambda = 4$, with a superdiagonal entry 1 reflecting the recurrence $(A - 4I)v_4 = v_3$.

Combining this with the simple eigenvalues $\lambda = 1$ and $\lambda = 2$, each contributing 1×1 blocks, we obtain:

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

The Jordan block for $\lambda = 4$ reflects the existence of a generalized eigenvector, and the entry 1 on the superdiagonal marks the action $Av_4 = 4v_4 + v_3$, in alignment with the structural logic developed in Section 4.6.

Step 4: Construct the Transformation Matrix

We now assemble the transformation matrix P using the (generalized) eigenvectors found in Step 2. The columns of P are ordered to reflect the Jordan structure:

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

More precisely:

- *The first two columns correspond to the true eigenvectors for $\lambda = 1$ and $\lambda = 2$.*
- *The last two columns form a Jordan chain for $\lambda = 4$, with v_3 followed by the generalized eigenvector v_4, satisfying $(A - 4I)v_4 = v_3$.*

This basis transforms A into its Jordan form:

$$P^{-1}AP = J,$$

preserving both the eigenstructure and the nilpotent behavior encoded in the Jordan block for $\lambda = 4$.

Step 5: Verify the Jordan Decomposition

To complete the decomposition, we verify that

$$A = PJP^{-1}.$$

This confirms that the transformation matrix P successfully conjugates A into its Jordan form J.

In particular, since the Jordan block for $\lambda = 4$ has size 2×2, this decomposition reflects that A is not diagonalizable. The presence of the superdiagonal 1 in the lower-right block arises precisely from the existence of a generalized eigenvector, as established in Step 2.

This example illustrates how Jordan Normal Form systematically captures non-diagonalizable structure. In particular, the eigenvalue $\lambda = 4$ lacks a full set of independent eigenvectors, leading to a Jordan block of size 2×2. The transformation matrix P encodes both eigenvectors and generalized eigenvectors, ensuring that $A = PJP^{-1}$ reflects the full spectral and nilpotent behavior.

Example 4.7.2 (Flexibility in Generalized Eigenvector Construction)**.** *We begin by defining the Jordan matrix J, structured as:*

$$
J = \begin{pmatrix}
4 & 1 & 0 & 0 & 0 & 0 \\
0 & 4 & 1 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 \\
0 & 0 & 0 & 5 & 1 & 0 \\
0 & 0 & 0 & 0 & 5 & 0 \\
0 & 0 & 0 & 0 & 0 & 3
\end{pmatrix}.
$$

The transformation matrix P is constructed in such a way that its columns correspond to the eigenvectors and generalized eigenvectors. Let P be:

$$
P = \begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 2 & 0 & 1 \\
1 & 1 & 1 & 0 & 6 & 1 \\
0 & 0 & 3 & 1 & 0 & 0 \\
2 & 0 & 2 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}.
$$

Using P and J, we compute

$$
A = PJP^{-1},
$$

which yields

$$
A = \begin{pmatrix}
-5 & 0 & 1 & -3 & 4 & -1 \\
101 & 13 & -8 & 37 & -51 & -2 \\
85 & 7 & -2 & 31 & -43 & -2 \\
50 & 4 & -4 & 22 & -25 & 0 \\
18 & 3 & -1 & 7 & -6 & -2 \\
86 & 7 & -7 & 31 & -43 & 3
\end{pmatrix}.
$$

In the previous example, the factorization of a non-diagonalizable matrix was explored in the direction $A \Rightarrow PJP^{-1}$. In this example, we consider the reverse process: starting from a Jordan matrix J and constructing a matrix A via $A = PJP^{-1}$. This confirms the bidirectional nature of the decomposition and illustrates how generalized eigenvectors enable reconstruction just as well as reduction.

Although the eigenvalues of a linear operator are uniquely determined (up to multiplicity), the choice of eigenvectors is not. Even among true eigenvectors—those satisfying $Tv = \lambda v$—there exists freedom when the eigenspace has dimension greater than one. In such cases, any basis for the eigenspace yields valid eigenvectors.

The flexibility becomes even greater when constructing generalized eigenvectors. These vectors arise from nilpotent components of T, and because nilpotent operators allow multiple distinct Jordan chains, we have several valid ways to build the transformation matrix P. As a result, while the Jordan form J is uniquely determined up to block ordering, the specific matrix P that satisfies $A = PJP^{-1}$ depends on the chosen bases of each (generalized) eigenspace.

In constructing a Jordan basis, one naturally wonders how much freedom exists in choosing the associated (generalized) eigenvectors.

To illustrate this, consider the case of the eigenvalue $\lambda = 4$ and $\lambda = 5$, corresponding to the 3×3 and 2×2 Jordan block:

$$J_4 = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix} \quad \textit{and} \quad J_5 = \begin{pmatrix} 5 & 1 \\ 0 & 5 \end{pmatrix}$$

Constructing Alternative Choices for Generalized Eigenvectors

We solve:

$$(A - 4I)v_2' = v_1.$$

Assume:

$$v_2' = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix},$$

Solving this system leads to:

$$v_2' = \begin{cases} x_1 = \frac{c}{2}, \\ x_2 = \frac{c+2}{2}, \\ x_3 = \frac{c+2}{2}, \\ x_4 = 0, \\ x_5 = c, \\ x_6 = 0. \end{cases}$$

Setting $c = 8$, we obtain:

$$v_2' = \begin{pmatrix} 4 \\ 5 \\ 5 \\ 0 \\ 8 \\ 0 \end{pmatrix}.$$

Now, using:

$$(A - 4I)v_3' = v_2'.$$

Solving similarly, we find:

$$v_3' = \begin{cases} x_1 = \frac{c-2}{2}, \\ x_2 = \frac{c+6}{2}, \\ x_3 = \frac{c+8}{2}, \\ x_4 = 3, \\ x_5 = c, \\ x_6 = 0. \end{cases}$$

Setting $c = 6$, we get:

$$v_3' = \begin{pmatrix} 2 \\ 6 \\ 7 \\ 3 \\ 6 \\ 0 \end{pmatrix}.$$

Furthermore, we can, in the same way, study the generalized eigenspace associated to $\lambda = 5$. We find, when we set

$$v_5' = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

, that the choice of direction becomes significantly variable, even though the size of this subspace is only of the type 2×2. Here the solution to the system

of equations is given by $(A - \lambda I)v_5' = v_4$.

$$v_5' = \begin{cases} x_1 = 1, \\ x_2 = 2c, \\ x_3 = 6, \\ x_4 = c, \\ x_5 = c + 1, \\ x_6 = c. \end{cases}$$

If we choose, for example, $c = 29$*, we will get*

$$v_5' = \begin{pmatrix} 1 \\ 58 \\ 6 \\ 29 \\ 30 \\ 29 \end{pmatrix}.$$

Final Transformation Matrix P'

With eigenvectors determined, the transformation matrix takes the form:

$$P' = \begin{pmatrix} 1 & 4 & 2 & 0 & 1 & 0 \\ 1 & 5 & 6 & 2 & 58 & 1 \\ 1 & 5 & 7 & 0 & 6 & 1 \\ 0 & 0 & 3 & 1 & 29 & 0 \\ 2 & 8 & 6 & 1 & 30 & 0 \\ 0 & 0 & 0 & 1 & 29 & 1 \end{pmatrix}.$$

Even though this transformation matrix P' ***differs*** *from the original choice of* P*, they both satisfy:*

$$A = P'JP'^{-1}.$$

Summary. *Jordan decomposition imposes strict structural constraints — the size and placement of Jordan blocks are determined by the operator. However, generalized eigenvectors offer freedom in construction: multiple chains can satisfy the same nilpotent relations. This flexibility allows for different transformation matrices* P*, while the Jordan matrix* J *remains unchanged.*

4.8 How Jordan Form Makes Functions of Non-Diagonalizable Matrices Computable

Introduction

We have seen, in 3.1, that when a matrix A is diagonalizable that is, $A = PDP^{-1}$ for some diagonal matrix D it becomes easy to compute expressions of the form $f(A)$ for any function f defined on the eigenvalues of A. In particular,

$$f(A) = Pf(D)P^{-1},$$

where $f(D)$ is just the diagonal matrix with $f(\lambda_i)$ along the diagonal.

But what happens if A is not diagonalizable? In that case, we can still write $A = PJP^{-1}$, where J is the Jordan Normal Form of A. If f is an analytic function, then it still holds that

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = \sum_{k=0}^{\infty} c_k PJ^k P^{-1} = P\left(\sum_{k=0}^{\infty} c_k J^k\right) P^{-1} = Pf(J)P^{-1}.$$

But $f(J)$ is no longer diagonal — it must be computed block by block, using a well-defined calculus of functions applied to Jordan blocks.

Remark. *For diagonalizable matrices, it is often sufficient that the function f is defined on the eigenvalues - even if f is not analytic - since only $f(\lambda)$ needs to be evaluated on the diagonal. However, for matrices that are not diagonalizable, and where the Jordan form is used, $f(J)$ needs to be defined block by block via a Taylor expansion around each eigenvalue. This requires that f is analytic in a neighborhood of each λ such that the derivatives $f^{(k)}(\lambda)$ exist. Therefore, non-analytic functions are generally not applicable to non-diagonalizable matrices.*

Investigating power functions of J

To understand how functions of a matrix behave when the matrix is not diagonalizable, we begin with the simplest possible Jordan block:

$$J = \begin{pmatrix} x & 1 \\ 0 & x \end{pmatrix}.$$

We compute J^2 directly:

$$J^2 = \begin{pmatrix} x & 1 \\ 0 & x \end{pmatrix}^2 = \begin{pmatrix} x^2 & 2x \\ 0 & x^2 \end{pmatrix}.$$

Notably, the entry in position (1,2) - the superdiagonal - is $2x$, the derivative of x^2. Is this a coincidence?

To investigate further, consider a larger Jordan block of size 6×6:

$$J_6 = \begin{pmatrix} x & 1 & 0 & 0 & 0 & 0 \\ 0 & x & 1 & 0 & 0 & 0 \\ 0 & 0 & x & 1 & 0 & 0 \\ 0 & 0 & 0 & x & 1 & 0 \\ 0 & 0 & 0 & 0 & x & 1 \\ 0 & 0 & 0 & 0 & 0 & x \end{pmatrix}.$$

Computing J_6^2, we obtain:

$$J_6^2 = \begin{pmatrix} x^2 & 2x & 1 & 0 & 0 & 0 \\ 0 & x^2 & 2x & 1 & 0 & 0 \\ 0 & 0 & x^2 & 2x & 1 & 0 \\ 0 & 0 & 0 & x^2 & 2x & 1 \\ 0 & 0 & 0 & 0 & x^2 & 2x \\ 0 & 0 & 0 & 0 & 0 & x^2 \end{pmatrix}.$$

Again, the first superdiagonal now consists of $2x$, and the second superdiagonal of 1. The structure persists and suggests a pattern: each superdiagonal entry corresponds to a derivative-like term of the diagonal entry x^2. What happens if we go one step further and examine J_6^3?

$$J_6^3 = \begin{pmatrix} x^3 & 3x^2 & 3x & 1 & 0 & 0 \\ 0 & x^3 & 3x^2 & 3x & 1 & 0 \\ 0 & 0 & x^3 & 3x^2 & 3x & 1 \\ 0 & 0 & 0 & x^3 & 3x^2 & 3x \\ 0 & 0 & 0 & 0 & x^3 & 3x^2 \\ 0 & 0 & 0 & 0 & 0 & x^3 \end{pmatrix}.$$

The pattern is now unmistakable: the entries above the diagonal resemble successive derivatives of the diagonal term x^3. Yet it's not merely derivative-like — their *precise coefficients* trace the rows of Pascal's triangle: 1, 3, 3, 1.

This observation reveals a deeper structure: computing powers of a Jordan block,

$$J^k = (xI + N)^k = \sum_{i=0}^{k} \binom{k}{i} x^{k-i} N^i,$$

follows the binomial expansion and naturally introduces the combinatorics of binomial coefficients. These coefficients reappear in the *Taylor expansion* of analytic functions applied to Jordan matrices:

$$f(J) = \sum_{i=0}^{n-1} \frac{f^{(i)}(x)}{i!} N^i.$$

Hence, the Pascal pattern is not incidental — it reflects both algebraic development and analytic regularity. The structure of $f(J)$ mirrors a truncated Taylor series centered at the eigenvalue x, with binomial coefficients acting as bridges between powers and derivatives.

To understand how $f(J)$ behaves, and why its structure reflects the analytic properties of f, we now turn to a detailed block-by-block expansion of $f(J)$.

Taylor Expansion of $f(J)$

Let f be an analytic function defined in a neighborhood of $\lambda \in \mathbb{C}$, and let $J = \lambda I + N \in \mathbb{C}^{n\times n}$ be a Jordan block, with N nilpotent such that $N^n = 0$. Since N is nilpotent, the Taylor expansion of f around λ yields:

$$f(J) = f(\lambda I + N) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\lambda)}{k!} (J - \lambda I)^k = \sum_{k=0}^{n-1} \frac{f^{(k)}(\lambda)}{k!} N^k.$$

Each term N^k contributes solely to the k-th superdiagonal of $f(J)$, and its coefficient is the k-th derivative of f evaluated at λ. This expansion is finite and fully determines $f(J)$ within the Jordan block.

Remark. *This formula shows that computing $f(J)$ reduces to evaluating derivatives of f at the eigenvalue λ, and placing them—scaled and normalized—on successive superdiagonals of the matrix. The binomial pattern observed earlier reflects this derivative-based structure.*

Example 4.8.1. *Let $f(x) = x^2 + 2x$. Then:*

$$f^{(0)}(x) = x^2 + 2x, \quad f^{(1)}(x) = 2x + 2, \quad f^{(2)}(x) = 2, \quad f^{(k)}(x) = 0 \textit{ for } k \geq 3.$$

Thus, for the Jordan block $J_6 = xI + N$, we obtain:

$$f(J_6) = f(x)I + f'(x)N + \frac{f''(x)}{2}N^2.$$

This yields:

$$f(J_6) = \begin{pmatrix} x^2+2x & 2x+2 & 1 & 0 & 0 & 0 \\ 0 & x^2+2x & 2x+2 & 1 & 0 & 0 \\ 0 & 0 & x^2+2x & 2x+2 & 1 & 0 \\ 0 & 0 & 0 & x^2+2x & 2x+2 & 1 \\ 0 & 0 & 0 & 0 & x^2+2x & 2x+2 \\ 0 & 0 & 0 & 0 & 0 & x^2+2x \end{pmatrix}$$

Example 4.8.2. *For $f(x) = e^x$, we obtain*

$$f(J) = e^\lambda \left(I + N + \frac{N^2}{2!} + \cdots + \frac{N^{n-1}}{(n-1)!} \right).$$

Since $D(e^\lambda) = e^\lambda$, in this case we can see e^λ as an unaffected factor on the right-hand side.

Hence, exponentials, polynomials, and trigonometric functions can all be computed directly on a Jordan block — and the familiar diagonal case is recovered when $N = 0$.

This result explains the structured appearance observed in the previous examples: the diagonal entries are $f(\lambda)$, the first superdiagonal is $f'(\lambda)$, the second is $\frac{f''(\lambda)}{2!}$, and so on. Each row reproduces a truncated Taylor series of f, but "shifted" along the upper triangle.

Since we have now established that e^J (and thus e^A) can be calculated without complications, we may turn to another Taylor series whose structure closely resembles that of the exponential: the sine function. Owing to this structural similarity, it is natural to expect that $\sin(J)$ (and therefore $\sin(A)$) are also well-defined. Extending the same logic, we might further argue that since e^J is welldefined, so too is $\log(I + J)$. With this line of reasoning, we now proceed to the next example.

Example 4.8.3 (Computing $\sin(J)$ and $\log(I + J)$ for a Jordan Block)**.** *Let us consider the Jordan block*

$$J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix},$$

which has nilpotent part

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{so that} \quad J = \lambda I + N.$$

(a) Sine function: *We consider the Maclaurin series for the sine function, defined as:*

$$\sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}.$$

If we now let $x = \lambda$ *and apply the general formula for* $f(J)$ *to the sine function given above for the Taylor expansion,* $f(J) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\lambda)}{k!} N^k$, *we get.*

$$\sin(J) = \sum_{j=0}^{2} \frac{\sin^{(j)}(\lambda)}{j!} N^j = \sum_{j=0}^{2} \frac{1}{j!} N^j \frac{d^j}{d\lambda^j} \left(\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} \right) =$$

$$= I \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} + N \frac{d}{d\lambda} \left(\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} \right) + \frac{N^2}{2} \frac{d^2}{d\lambda^2} \left(\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} \right) =$$

$$= I \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} + N \sum_{k=0}^{\infty} \frac{(-1)^k}{2k!} \lambda^{2k} + \frac{N^2}{2} \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k-1)!} \lambda^{2k-1} = *$$

If we put $k = j+1$ *in the last term, we get* $\sum_{k=1}^{\infty} \frac{(-1)^k}{(2k-1)!} \lambda^{2k-1} = \sum_{j=0}^{\infty} \frac{(-1)^{j+1}}{(2(j+1)-1)!} \lambda^{2(j+1)-1} = -\sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} \lambda^{2j+1}$. *There is now nothing to prevent setting* $j = k$, *which gives* $\sum_{k=1}^{\infty} \frac{(-1)^k}{(2k-1)!} \lambda^{2k-1} = -\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1}$

$$* = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} I + \sum_{k=0}^{\infty} \frac{(-1)^k}{2k!} \lambda^{2k} N - \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \lambda^{2k+1} \frac{N^2}{2} =$$

$$= \sin(\lambda) I + \cos(\lambda) N - \sin(\lambda) \frac{N^2}{2}$$

This series defines a matrix whose:
- *diagonal entries are* $\sin(\lambda)$,
- *first superdiagonal is* $\cos(\lambda)$,
- *second superdiagonal is* $-\frac{1}{2}\sin(\lambda)$.

Thus,

$$\sin(J) = \begin{pmatrix} \sin(\lambda) & \cos(\lambda) & -\frac{1}{2}\sin(\lambda) \\ 0 & \sin(\lambda) & \cos(\lambda) \\ 0 & 0 & \sin(\lambda) \end{pmatrix}.$$

This example shows that $f(J)$ is upper triangular with entries constructed from the derivatives of f evaluated at λ, positioned along the corresponding superdiagonals.

Remark. *In the expression*

$$\sin(J) = \sin(\lambda)I + \cos(\lambda)N - \tfrac{1}{2}\sin(\lambda)N^2,$$

each term corresponds to the truncated Taylor series of $\sin(x)$*, evaluated at λ and positioned along successive superdiagonals. The series components can be regrouped as nested expansions, revealing how analytic structure filters into matrix form. This perspective connects directly to Example 4.8.1 and makes the layered appearance of $f(J)$ fully transparent.*

(b) Function of the natural logarithm $\log(I + J)$***:***
Let $f(x) = \log(1 + x)$, and set $x = \lambda$. Then apply the general expansion

$$f(J) = \sum_{k=0}^{2} \frac{f^{(k)}(\lambda)}{k!} N^k$$

Since

$$f'(\lambda) = \frac{1}{1+\lambda} \quad \textit{and} \quad f''(\lambda) = -\frac{1}{(1+\lambda)^2},$$

we obtain

$$\log(I + J) = \log(1 + \lambda) \cdot I + \frac{1}{1+\lambda} \cdot N - \frac{1}{(1+\lambda)^2} \cdot \frac{N^2}{2}.$$

Written explicitly, the matrix representation becomes:

$$\log(I + J) = \begin{pmatrix} \log(1+\lambda) & \frac{1}{1+\lambda} & -\frac{1}{2(1+\lambda)^2} \\ 0 & \log(1+\lambda) & \frac{1}{1+\lambda} \\ 0 & 0 & \log(1+\lambda) \end{pmatrix}.$$

Remark. Each term corresponds to the Taylor expansion of $\log(1 + x)$ *evaluated at $x = \lambda$, with derivatives positioned along successive superdiagonals. This result confirms that* $\log(I + J)$ *is well-defined as long as N is nilpotent, allowing the analytic expansion to truncate naturally.*

Square Roots and Pascal Patterns

The triangular derivative pattern we observed in $f(J)$ applies not only to standard analytic functions like e^x, $\sin(x)$ and $\log(1+x)$, but also to square roots. Even though $f(x) = \sqrt{x}$ involves a non-polynomial function, its Taylor expansion around $\lambda > 0$ retains the same binomial logic — and yields a matrix representation with entries drawn from derivatives and Pascal's triangle.

Before proceeding, it is worth noting a general principle: if a certain function is well-defined in a given context, there is reason to believe that its inverse may also be well-defined. We previously examined power functions on a Jordan block J with eigenvalue $\lambda = x$, such as J^2, which revealed a structure explained naturally by the binomial expansion of $(\lambda I + N)^2$.

This observation suggests that even non-polynomial operations could follow a similar pattern when applied to Jordan blocks. With this in mind, we now turn to a particularly interesting case — the square root of a Jordan block. What does $\sqrt{J}$ look like? This will be the focus of the next example.

Example 4.8.4 (Computing $\sqrt{J}$ for a Jordan block)**.** *Let*

$$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = \lambda I + N, \quad N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad N^2 = 0.$$

Let $f(x) = \sqrt{x}$. Then:

$$f(J) = \sqrt{\lambda} I + \frac{f'(\lambda)}{1!} N = \begin{pmatrix} \sqrt{\lambda} & \frac{f'(\lambda)}{1!} \\ 0 & \sqrt{\lambda} \end{pmatrix}$$

Since $f'(x) = \frac{1}{2\sqrt{x}}$, we obtain:

$$\sqrt{J} = \begin{pmatrix} \sqrt{\lambda} & \frac{1}{2\sqrt{\lambda}} \\ 0 & \sqrt{\lambda} \end{pmatrix}$$

Verification: $(\sqrt{J})^2 = J$

This shows that the square root of a Jordan block respects the binomial derivative structure observed earlier — with entries matching the truncated Taylor expansion of $\sqrt{x}$.

Remark. *This example illustrates Theorem 4.4.3 and shows that square roots of operators can be constructed explicitly using Jordan blocks, even when diagonalization fails.*

Summary. Jordan Normal Form provides a structural foundation for computing matrix functions — even when diagonalization fails. The formula $f(A) = Pf(J)P^{-1}$ remains valid, with $f(J)$ constructed from derivatives of f evaluated at eigenvalues. The triangular structure of $f(J)$ reflects Pascal's triangle, appearing naturally in powers, exponentials, and Taylor expansions. As shown by the square root example, even non-polynomial functions can be applied blockwise — confirming that canonical structure enables analytic flexibility.

Transition to Chapter 5. Having reached the final section of our exploration, we now pause to reflect on what the Jordan Normal Form ultimately reveals — not just about matrices and linear operators, but about the broader algebraic structures at play. The next chapter ties together the mathematical journey that began with eigenvalues and concludes with functions of non-diagonalizable matrices.

5 Concluding Reflections

This thesis has explored the progression from classical eigenvalue theory to the canonical power of the Jordan Normal Form. What began as the study of eigenvalues and invariant subspaces developed into a deeper understanding of linear operators — their algebraic constraints, geometric manifestations, and structural possibilities.

The role of diagonalization has been central: it offers a pathway to simplify operator behavior, but only when sufficient eigenvectors exist. In cases where diagonalization fails, Jordan Normal Form extends the framework by introducing nilpotent structure and generalized eigenvectors.

Earlier, Chapter 3 showed how diagonalization enabled the application of functions via $A = PDP^{-1} \Rightarrow f(A) = Pf(D)P^{-1}$, emphasizing how functions of matrices rely heavily on eigenstructure. But limitations emerged — not all operators are diagonalizable. Chapter 4 responded by unfolding the machinery of Jordan Normal Form, and Section 4.8 revealed a crucial insight: if a function f is analytic, the expression $A = PJP^{-1} \Rightarrow f(A) = Pf(J)P^{-1}$ remains valid. This allows for functional evaluation — such as computing expressions like e^A or $\sin(A)$ — even when A lacks sufficient eigenvectors.

This insight elegantly closes the conceptual circle. The computational advantages first attributed to diagonalizability — enabling the functional evaluation of matrices — are preserved and extended through Jordan theory and analytic expansion. The decomposition $A = PJP^{-1}$ is not merely symbolic; it's a powerful operational tool that offers structural clarity and computational access even in the absence of diagonalization.

While all results followed Axler's formal structure, the examples and constructions have illustrated how abstract theory becomes concrete. It is this interplay between structure and flexibility, rigidity and freedom, that gives Jordan Normal Form its lasting mathematical significance. Where diagonalization ends, Jordan theory begins — and with it, a deeper insight into the algebra that governs transformations.

References

[1] Sheldon Axler. *Linear Algebra Done Right.* Springer, 4th edition, 2024.