The Companion to BMP

Contents

1 Algebraic functions

1.1 The Carmichael function

For a positive integer n, the *Carmichael function* $\lambda(n)$ is defined to be the exponent of the multiplicative group

$$(\mathbb{Z}/n\mathbb{Z})^{\times}.$$

Equivalently,

$$\lambda(n) = \min\{k \geqslant 1 : a^k \equiv 1 \pmod{n} \text{ for all } a \text{ with } \gcd(a,n) = 1\}.$$

Thus $\lambda(n)$ is the smallest positive integer such that

$$a^{\lambda(n)} \equiv 1 \pmod{n} \quad \text{for every unit } a \in (\mathbb{Z}/n\mathbb{Z})^{\times}.$$

Relation to Euler's totient function

Euler's theorem states that

$$a^{\varphi(n)} \equiv 1 \pmod{n} \quad \text{for all } \gcd(a,n) = 1.$$

Hence $\lambda(n) \mid \varphi(n)$, but in general

$$\lambda(n) \leqslant \varphi(n),$$

and $\lambda(n)$ is often strictly smaller.

In contrast to $\varphi(n)$, which gives a *sufficient* exponent, the Carmichael function gives the *minimal universal exponent* valid for all units modulo n.

Values on prime powers

For a prime p and integer $k \geqslant 1$,

$$\lambda(p^k) = \begin{cases} p^{k-1}(p-1), & \text{if } p \text{ is odd,} \\ 1, & p = 2,\ k = 1, \\ 2, & p = 2,\ k = 2, \\ 2^{k-2}, & p = 2,\ k \geqslant 3. \end{cases}$$

In particular, for odd primes,

$$\lambda(p^k) = \varphi(p^k).$$

General formula

If

$$n = \prod_{i=1}^{r} p_i^{k_i}$$

is the prime factorization of n, then

$$\boxed{\lambda(n) = \operatorname{lcm}\big(\lambda(p_1^{k_1}), \dots, \lambda(p_r^{k_r})\big).}$$

This follows from the Chinese Remainder Theorem and the fact that the exponent of a finite abelian group is the least common multiple of the exponents of its direct factors.

Example: $n = 2025$

We have

$$2025 = 3^4 \cdot 5^2.$$

Since both primes are odd,

$$\lambda(3^4) = \varphi(3^4) = 3^3(3-1) = 54, \qquad \lambda(5^2) = \varphi(5^2) = 5(5-1) = 20.$$

Therefore,

$$\lambda(2025) = \operatorname{lcm}(54, 20) = 540.$$

Why the Carmichael function appears in congruence problems

Suppose we require a congruence of the form

$$x^m \equiv x^n \pmod{n} \quad \text{for all integers } x.$$

For integers x with $\gcd(x, n) = 1$, this is equivalent to

$$x^{n-m} \equiv 1 \pmod{n} \quad \text{for all units } x.$$

By definition of $\lambda(n)$, this holds if and only if

$$\lambda(n) \mid (n - m).$$

Thus the Carmichael function gives the *sharp necessary and sufficient condition* on the exponent difference $n - m$ in such problems.

Summary

- $\lambda(n)$ is the smallest exponent that annihilates all units modulo n.
- Always $\lambda(n) \mid \varphi(n)$.
- For odd prime powers, $\lambda(p^k) = \varphi(p^k)$.
- In exponent congruence problems, $\lambda(n)$ gives the minimal condition required for all coprime residues.

1.2 Euler's Totient Function (ϕ)

The function ϕ used in the solution is **Euler's totient function** (also known as the Euler phi function). It is a fundamental arithmetic function in number theory.

1. Definition

For a positive integer n, $\phi(n)$ counts the number of integers between 1 and n that are **relatively prime** (coprime) to n. Two numbers are relatively prime if their greatest common divisor (GCD) is 1.

$$\phi(n) = |\{k \in \{1, \ldots, n\} : \gcd(k, n) = 1\}|$$

Example: $n = 10$ The integers from 1 to 10 are: $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.

- Numbers sharing a factor with 10 (not coprime): $2, 4, 5, 6, 8, 10$.
- Numbers relatively prime to 10: **1, 3, 7, 9**.

Since there are 4 such numbers, $\phi(10) = 4$.

2. Calculation Formula

The standard formula relies on the prime factorization of n. If the prime factorization of n is given by $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, then:

$$\phi(n) = n \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_r}\right)$$

Special Case: Prime Powers For a single prime power p^k, the formula simplifies to:

$$\phi(p^k) = p^k - p^{k-1}$$

This is because there are p^k total numbers, and the multiples of p ($1p, 2p, \ldots, p^{k-1}p$) must be subtracted. There are exactly p^{k-1} such multiples.

Applying this to the problem factors:

- For $3^4 = 81$:
$$\phi(81) = 3^4 - 3^3 = 81 - 27 = \mathbf{54}$$
- For $5^2 = 25$:
$$\phi(25) = 5^2 - 5^1 = 25 - 5 = \mathbf{20}$$

3. Why is it used? (Euler's Theorem)

We used this function because of **Euler's Theorem**, which generalizes Fermat's Little Theorem. It states that if $\gcd(a, n) = 1$, then:

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

This implies that the powers of x modulo n repeat in a cycle of length $\phi(n)$.

- If x is coprime to the modulus, $x^{k+\phi(n)} \equiv x^k \pmod{n}$.
- This is why, when solving $x^n \equiv x^m$, we required the difference $n - m$ to be a multiple of $\phi(p^k)$.

1.3 Falling factorial

1.4 Double factorial

1.5 q-factorial

1.6 Gaussian binomial coefficient

2 Algebra

2.1 $\mathbb{Z}_p \cong \mathbb{F}_p$

3 Linear Algebra

3.1 Generalized Eigenspace

In linear algebra, a generalized eigenspace is an expansion of the traditional eigenspace. It is necessary because not every matrix has enough "ordinary" eigenvectors to form a basis (a property known as being *defective*).

3.1.1 Formal Definition

Let V be a finite-dimensional vector space over a field $\mathbb{F}$ (typically $\mathbb{C}$), and let $T : V \to V$ be a linear operator. Let λ be an eigenvalue of T with algebraic multiplicity m.

The **generalized eigenspace** of T associated with λ, denoted by G_λ, is the set of all vectors $v \in V$ such that

$$(T - \lambda I)^k v = 0$$

for some positive integer $k \geqslant 1$.

While the definition says "for some k," we don't need to look forever. For a vector space of dimension n, if a vector is going to be "annihilated" by the operator $(T - \lambda I)$, it will definitely happen by the time $k = n$, that is, in a vector space of dimension n, it can be shown that if such a k exists, then $k \leqslant n$. Thus, the generalized eigenspace can be expressed as the kernel of the n-th power of the operator:

$$G_\lambda = \ker(T - \lambda I)^n$$

Why do we need them? The ordinary eigenspace $E_\lambda = \ker(T - \lambda I)$ only contains vectors that T scales exactly by λ. However, many matrices (like those with Jordan blocks larger than 1×1) are missing eigenvectors.

3.1.2 Key Properties

1. **Subspace Property:** G_λ is a subspace of V and is invariant under T (i.e., $T(G_\lambda) \subseteq G_\lambda$).

2. **Inclusion:** The ordinary eigenspace $E_\lambda = \ker(T - \lambda I)$ is a subset of the generalized eigenspace G_λ.

3. **Dimension:** The dimension of G_λ is equal to the algebraic multiplicity m of the eigenvalue λ.

4. **Primary Decomposition:** If the characteristic polynomial of T splits completely over $\mathbb{F}$, then V is the direct sum of the generalized eigenspaces:

$$V = G_{\lambda_1} \oplus G_{\lambda_2} \oplus \cdots \oplus G_{\lambda_m}$$

3.1.3 Structure of Generalized Eigenvectors

A vector $v \in G_\lambda$ is said to be a **generalized eigenvector of rank** k if

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

These vectors form the "Jordan chains" that allow for the construction of the Jordan Canonical Form when T is not diagonalizable.

Visualizing the "Chain". Generalized eigenvectors often exist in a "chain" starting from a rank-1 eigenvector:

- $(T - \lambda I)v_1 = 0$ (Ordinary eigenvector)
- $(T - \lambda I)v_2 = v_1$ (Rank-2 generalized eigenvector)
- $(T - \lambda I)v_3 = v_2$ (Rank-3 generalized eigenvector)

Multiplying v_3 by $(T - \lambda I)$ three times will eventually hit zero, even though the first two multiplications did not.

To provide a concrete example, we will find the generalized eigenspace for a 3×3 matrix that is not diagonalizable.

3.1.4 Worked Example

Consider the matrix T representing a single 3×3 Jordan block:

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

The characteristic polynomial is $p(\lambda) = (2 - \lambda)^3$, yielding the eigenvalue $\lambda = 2$ with algebraic multiplicity 3.

1. Ordinary Eigenspace

The ordinary eigenspace $E_2 = \ker(T - 2I)$ is found by solving:

$$(T - 2I)\mathbf{v} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This yields $y = 0$ and $z = 0$, so $E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$. Since $\dim(E_2) = 1 < 3$, T is not diagonalizable.

2. Generalized Eigenspace

We compute higher powers of $(T - 2I)$:

$$(T - 2I)^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (T - 2I)^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The generalized eigenspace is $G_2 = \ker(T - 2I)^3 = \mathbb{C}^3$.

3. Jordan Chain

We can pick a vector $\mathbf{v}_3 = (0, 0, 1)^T$. Then the chain is:

$$\begin{aligned} \mathbf{v}_3 &= (0, 0, 1)^T \quad \text{(Rank 3)} \\ \mathbf{v}_2 &= (T - 2I)\mathbf{v}_3 = (0, 1, 0)^T \quad \text{(Rank 2)} \\ \mathbf{v}_1 &= (T - 2I)\mathbf{v}_2 = (1, 0, 0)^T \quad \text{(Rank 1, Eigenvector)} \end{aligned}$$

3.2 Simultaneous Derivation of Rank-1 Identities via Schur Complements

Introduction

The **Matrix Determinant Lemma** and the **Sherman–Morrison Formula** are two fundamental identities in linear algebra regarding rank-1 updates to a matrix. Both can be derived simultaneously by analyzing a specific block matrix and its Schur complements.

The Block Matrix Construction

Consider the $(n + 1) \times (n + 1)$ block matrix M:

$$M = \begin{pmatrix} A & u \\ v^T & -1 \end{pmatrix}$$

where $A \in \mathbb{R}^{n\times n}$ is an invertible matrix, and $u, v \in \mathbb{R}^n$ are column vectors.

1. Deriving the Matrix Determinant Lemma

The determinant of a block matrix can be expressed using the Schur complement of either the bottom-right or top-left block.

Method A: Schur complement of the block -1

Using the formula $\det(M) = \det(S)\det(P - QS^{-1}R)$:

$$\det(M) = \det(-1)\det(A - u(-1)^{-1}v^T)$$

$$\det(M) = -1 \cdot \det(A + uv^T)$$

Method B: Schur complement of the block A

Using the formula $\det(M) = \det(P)\det(S - RP^{-1}Q)$:

$$\det(M) = \det(A)\det(-1 - v^T A^{-1}u)$$

Since the second term is a scalar, we can write:

$$\det(M) = -\det(A)(1 + v^T A^{-1}u)$$

Conclusion

Equating Method A and Method B:

$$-\det(A + uv^T) = -\det(A)(1 + v^T A^{-1}u)$$

Multiplying by -1 gives the **Matrix Determinant Lemma**:

$$\det(A + uv^T) = \det(A)(1 + v^T A^{-1}u)$$

2. Deriving the Sherman–Morrison Formula

We now consider the inverse of the block matrix M. There are two ways to express the top-left $n \times n$ block of M^{-1}.

Direct Inversion via the Schur Complement of -1

From the general formula for block inversion, the top-left block is the inverse of the Schur complement of the bottom-right block:

$$(M^{-1})_{11} = (A - u(-1)^{-1}v^T)^{-1} = (A + uv^T)^{-1}$$

Inversion via the Schur Complement of A

Alternatively, the block inversion formula expressed in terms of A^{-1} gives the top-left block as:

$$(M^{-1})_{11} = A^{-1} + A^{-1}u(-1 - v^T A^{-1}u)^{-1}v^T A^{-1}$$

Since $(-1 - v^T A^{-1}u)$ is a scalar, we can move it to the denominator:

$$(M^{-1})_{11} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u}$$

Conclusion

Equating these two expressions for the $(1, 1)$ block yields the **Sherman–Morrison Formula**:

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u}$$

4 Combinatorics

The Hockey-stick Identity

The combinatorial identity

$$\sum_{i=r}^{n} \binom{i}{r} = \binom{n+1}{r+1}$$

is affectionately known as the **Hockey-stick Identity** (or the *Christmas Stocking Identity*).

Why the Name?

The name comes from the geometric shape the terms form when highlighted on Pascal's Triangle.

- The terms being summed form a diagonal line (the "shaft" of the stick).
- The result of the sum is located diagonally below the last term, creating a sharp turn (the "blade" of the stick).

Visualizing the "Stick"

Consider the example for $r = 2$ and $n = 5$:

$$\binom{\mathbf{2}}{\mathbf{2}} + \binom{\mathbf{3}}{\mathbf{2}} + \binom{\mathbf{4}}{\mathbf{2}} + \binom{\mathbf{5}}{\mathbf{2}} = \underline{\binom{\mathbf{6}}{\mathbf{3}}}$$

In numbers: $1 + 3 + 6 + 10 = 20$.

Below is a representation of Pascal's Triangle. The bolded numbers form the shaft, and the underlined number is the blade (the sum).

$$
\begin{array}{ccccccccccccc}
 & & & & & 1 & & & & & & & \\
 & & & & 1 & & 1 & & & & & & \\
 & & & \mathbf{1} & & 2 & & 1 & & & & & \\
 & & 1 & & \mathbf{3} & & 3 & & 1 & & & & \\
 & 1 & & 4 & & \mathbf{6} & & 4 & & 1 & & & \\
1 & & 5 & & 10 & & \mathbf{10} & & 5 & & 1 & & \\
\hline
1 & & 6 & & 15 & & \underline{\mathbf{20}} & & 15 & & 6 & & 1
\end{array}
$$

Notice how the bold numbers descend diagonally, and the final sum steps down and to the right, resembling the shape of a hockey stick.

Discrete Calculus and Falling Factorials

The term "Falling Factorial" refers to a specific basis for polynomials that simplifies summation in the same way that standard powers simplify integration.

1. Definition

The falling factorial of x of degree r, denoted as $x^{\underline{r}}$, is defined as the product of r terms starting at x and decreasing by 1:

$$x^{\underline{r}} = x(x-1)(x-2)\ldots(x-r+1)$$

This relates directly to the binomial coefficient:

$$\binom{x}{r} = \frac{x^{\underline{r}}}{r!} \implies x^{\underline{r}} = r!\binom{x}{r}$$

2. The Discrete Derivative

In standard calculus, the derivative is $\frac{d}{dx}$. In discrete calculus, the equivalent operator is the forward difference operator Δ:

$$\Delta f(x) = f(x+1) - f(x)$$

The falling factorials satisfy a **Power Rule** for the difference operator that is nearly identical to the power rule for derivatives ($\frac{d}{dx}x^n = nx^{n-1}$):

$$
\begin{aligned}
\Delta x^{\underline{r}} &= (x+1)^{\underline{r}} - x^{\underline{r}} \\
&= (x+1)x(x-1)\ldots(x-r+2) - x(x-1)\ldots(x-r+1) \\
&= x(x-1)\ldots(x-r+2)\left[(x+1)-(x-r+1)\right] \\
&= x^{\underline{r-1}} \cdot [r] \\
&= rx^{\underline{r-1}}
\end{aligned}
$$

3. The Summation Rule (Discrete Integration)

Just as integration is the inverse of differentiation, summation is the inverse of the difference operator.

$$\text{If } \Delta F(k) = f(k), \text{ then } \sum_{k=a}^{b-1} f(k) = F(b) - F(a)$$

Applying this to our Power Rule yields the **Discrete Power Rule for Summation**:

$$\sum_{k=0}^{n-1} k^{\underline{r}} = \frac{n^{\underline{r+1}}}{r+1}$$

This is structurally identical to the integral power rule $\int_0^n x^r \, dx = \frac{n^{r+1}}{r+1}$.

4. Conclusion

Since any polynomial $P(k)$ can be written as a linear combination of falling factorials (or binomial coefficients), we can sum the polynomial term-by-term using this simple rule, producing a result that is a polynomial of degree one higher.

5 Young diagrams

5.1 Partitions of Integers

The relationship between the **partitions of an integer** and **Young diagrams** (also known as Ferrers diagrams) is a fundamental correspondence in combinatorics and representation theory. A Young diagram serves as the visual and geometric realization of an integer partition.

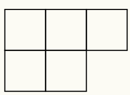
5.1.1 Formal Definition

A partition of a positive integer n is a way of writing n as a sum of positive integers where the order of the summands is irrelevant. Traditionally, partitions are written in non-increasing order:

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k) \quad \text{where} \quad \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k > 0 \quad \text{and} \quad \sum_{i=1}^{k} \lambda_i = n$$

A **Young diagram** associated with the partition λ is a finite collection of boxes, or cells, arranged in left-justified rows, where the i-th row contains λ_i boxes.

Example: For $n = 5$, the partition $\lambda = (3, 2)$ is represented by a diagram with 3 boxes in the first row and 2 boxes in the second row:



5.1.2 Conjugate Partitions

One of the most powerful features of the Young diagram is the definition of the **conjugate partition**, denoted λ'. The conjugate is obtained by reflecting the Young diagram across its main diagonal (essentially swapping rows for columns).

For the partition $\lambda = (3, 2)$, the reflection yields:

- Column 1 becomes Row 1: 2 boxes.
- Column 2 becomes Row 2: 2 boxes.
- Column 3 becomes Row 3: 1 box.

Thus, the conjugate partition is $\lambda' = (2, 2, 1)$.

5.1.3 Mathematical Applications

Young diagrams allow for the study of complex algebraic structures through visual combinatorics:

Concept	Relationship to Young Diagrams
Symmetric Group S_n	The irreducible representations of S_n are indexed by the Young diagrams of size n.
Hook Length Formula	The dimension of an irreducible representation is calculated by filling the boxes of a Young diagram with their respective "hook lengths."
Young Tableaux	Filling a Young diagram with integers (standard or semi-standard) is used to study Schur functions and polynomial rings.

5.1.4 Summary

If partitions represent the **arithmetic** of decomposing an integer, Young diagrams represent the **geometry**. They reveal symmetries, such as conjugation and the dominance order, that are not immediately apparent from numerical lists.

5.2 Young Diagrams and the Jordan Canonical Form

5.2.1 The Relationship Between JCF and Young Diagrams

The structure of a linear operator T on a generalized eigenspace is uniquely determined by the sizes of its Jordan blocks. This structure corresponds directly to a Young diagram.

The relationship between Young tableaux (specifically Young diagrams) and the Jordan Canonical Form (JCF) is found in how we track the structure of nilpotent operators. Essentially, the Young diagram is a geometric map of the sizes of the Jordan blocks associated with a single eigenvalue.

5.2.2 The Partition of the Generalized Eigenspace

For a given eigenvalue λ of an $n \times n$ matrix, the algebraic multiplicity m tells us the total number of boxes in the Young diagram associated with λ.

If the Jordan blocks associated with λ have sizes $k_1, k_2, \ldots, k_t$, then these sizes form a partition of m:

$$k_1 + k_2 + \cdots + k_t = m$$

By convention, we arrange these in non-increasing order. This partition is represented by a Young diagram where each row corresponds to a Jordan block.

Example: If an eigenvalue λ has three Jordan blocks of sizes 3, 2, and 1, the Young diagram is:

5.2.3 The "Weyr" Characteristic and Columns

The relationship becomes even deeper when you look at the columns of the Young diagram. This is related to the Weyr Canonical Form, which is a reordering of the Jordan form.

- Row lengths: Tell you the sizes of the individual Jordan blocks (k_i).
- Column lengths: Tell you the "jumps" in the dimensions of the kernels of powers of $(A - \lambda I)$.

If $n_i = \dim(\ker(A - \lambda I)^i)$, then the number of boxes in the i-th column of the Young diagram is:

$$c_i = n_i - n_{i-1}$$

This tells us how many Jordan blocks have a size of at least i.

5.2.4 Visualizing the Generalized Eigenspace

If you pick a basis for the generalized eigenspace consisting of Jordan chains, the Young diagram effectively "slots" these vectors into boxes:

- Each row of the diagram represents a single Jordan chain (a single Jordan block).
- The first column represents the "true" eigenvectors (the basis for the eigenspace $\ker(A - \lambda I)$).
- The subsequent columns represent the generalized eigenvectors of higher ranks.

5.2.5 Row and Column Correspondence

For a single eigenvalue λ, the Jordan blocks provide a partition of the algebraic multiplicity m.

- **Row Lengths:** Each row in the Young diagram corresponds to the size of a single Jordan block.

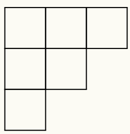
- **Geometric Multiplicity:** The total number of rows equals the number of Jordan blocks (the dimension of the eigenspace $\ker(T - \lambda I)$).
- **Column Lengths:** The height of the k-th column tells us how many Jordan blocks have size at least k.

Feature of Jordan Form	**Feature of Young Diagram**
Algebraic Multiplicity	Total number of boxes (n)
Geometric Multiplicity	Number of rows
Size of Largest Block	Number of columns
Jordan Chain Lengths	Lengths of the individual rows
$\dim(\ker(A - \lambda I)^k) - \dim(\ker(A - \lambda I)^{k-1})$	Length of the k-th column

5.2.6 The "Nilpotent" Case

Let N be a nilpotent matrix. The similarity class of N is uniquely determined by a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ of n. The corresponding Young diagram has rows of length λ_i, representing Jordan blocks $J_{\lambda_i}(0)$.

For example, a 6×6 nilpotent matrix with blocks of size 3, 2, and 1 corresponds to:



The geometric multiplicity (nullity of N) is the number of rows (3), and the index of nilpotency is the number of columns (3).

5.2.7 Worked Example: Kernel Dimensions to Jordan Blocks

To determine the Young diagram (and thus the Jordan Form) of a nilpotent operator N, we analyze the sequence of nullities $n_k = \dim(\ker N^k)$.

5.2.8 Example Scenario

Consider a 4×4 nilpotent matrix N with the following nullities:

$$n_1 = \dim(\ker N) = 2$$
$$n_2 = \dim(\ker N^2) = 3$$
$$n_3 = \dim(\ker N^3) = 4$$

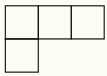
5.2.9 Step 1: Calculate Column Lengths

The number of boxes in the i-th column, c_i, is the growth in nullity: $c_i = n_i - n_{i-1}$.

- $c_1 = n_1 - n_0 = 2 - 0 = 2$
- $c_2 = n_2 - n_1 = 3 - 2 = 1$
- $c_3 = n_3 - n_2 = 4 - 3 = 1$

5.2.10 Step 2: The Resulting Diagram and Partition

By assembling these columns, we obtain the following Young diagram:



Reading the **rows** gives the partition $\lambda = (3, 1)$. This implies the matrix consists of one 3×3 Jordan block and one 1×1 Jordan block:

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

Nilpotent similarity classes via Young diagrams – Part A

Theorem: Two nilpotent $n \times n$ matrices A and B are similar if and only if they are associated with the same Young diagram.

Proof. Let A be a nilpotent operator. By the Jordan Canonical Form theorem, A is similar to a direct sum of Jordan blocks of sizes $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k$. We associate A with a Young diagram Y_A having row lengths λ_i.

1. The Invariants Similarity preserves the dimensions of subspaces derived from the operator. Specifically, define the sequence:

$$d_j(A) = \dim(\ker(A^j))$$

If $A \sim B$, then $d_j(A) = d_j(B)$ for all $j \geqslant 1$.

2. Relating Kernels to Columns Consider the basis vectors of the Jordan blocks placed into the boxes of Y_A. The operator A maps a vector in box (r, c) to the box $(r, c - 1)$, and maps column 1 to zero. Therefore, A^j annihilates exactly those basis vectors residing in the first j columns of the diagram.

$$\dim(\ker(A^j)) = \text{Total number of boxes in the first } j \text{ columns of } Y_A$$

3. Determining the Shape Let C_j be the length of the j-th column of Y_A (which corresponds to the number of Jordan blocks of size $\geqslant j$). From the previous step:

$$d_j(A) - d_{j-1}(A) = C_j$$

Thus, the geometric shape of the diagram (its column lengths) is uniquely determined by the similarity invariants $\{d_j\}$.

Conclusion

- If $A \sim B$, then $d_j(A) = d_j(B)$, so $C_j(A) = C_j(B)$. The diagrams are identical.
- If $Y_A = Y_B$, the Jordan blocks have the same sizes. Thus A and B share the same Jordan Normal Form (up to permutation), implying $A \sim B$.

□

Nilpotent similarity classes via Young diagrams – Part B

Let F be an algebraically closed field (or just a field over which the Jordan form exists). An $n \times n$ matrix N is *nilpotent* if $N^m = 0$ for some $m \geqslant 1$, equivalently if all eigenvalues are 0. It is a corollary of Jordan theory that nilpotent similarity classes are classified by the multiset of Jordan block sizes. Here is a proof phrased in terms of partitions and Young diagrams.

1. From a nilpotent operator to a partition

Let V be an n-dimensional vector space over F, and let $N : V \to V$ be nilpotent. For $r \geqslant 0$ set

$$k_r := \dim\ker(N^r), \qquad k_0 = \dim\ker(I) = 0.$$

Then $0 = k_0 \leqslant k_1 \leqslant k_2 \leqslant \cdots \leqslant n$, and for r large enough $k_r = n$. Define the increments

$$a_r := k_r - k_{r-1} \qquad (r \geqslant 1).$$

Clearly $a_r \geqslant 0$ and $\sum_{r \geqslant 1} a_r = n$. The next lemma shows $(a_1, a_2, \ldots)$ is a *partition* of n, i.e. a weakly decreasing sequence of nonnegative integers with sum n.

Lemma 1. *For all $r \geqslant 1$, one has $a_r \geqslant a_{r+1}$. Equivalently,*

$$k_r - k_{r-1} \;\geqslant\; k_{r+1} - k_r.$$

Proof. Consider the linear map induced by N:

$$\overline{N} :\; V/\ker(N^r) \longrightarrow V/\ker(N^{r-1}), \qquad \overline{N}([v]) = [Nv].$$

It is well-defined because if $v - v' \in \ker(N^r)$, then $N(v - v') \in \ker(N^{r-1})$. Moreover, $\overline{N}$ is injective: if $[Nv] = 0$ in $V/\ker(N^{r-1})$, then $Nv \in \ker(N^{r-1})$, hence $N^r v = 0$, so $[v] = 0$ in $V/\ker(N^r)$. Thus

$$\dim(V/\ker(N^r)) \leqslant \dim(V/\ker(N^{r-1})),$$

i.e. $n - k_r \leqslant n - k_{r-1}$. Rewriting gives $k_r - k_{r-1} \geqslant k_{r+1} - k_r$, which is $a_r \geqslant a_{r+1}$. □

Hence $\lambda'(N) := (a_1, a_2, \ldots)$ is a partition of n. We draw its *Young diagram*: a_1 boxes in the first row, a_2 in the second row, etc.

Similarity invariance. If N and M are similar, then $\dim\ker(N^r) = \dim\ker(M^r)$ for all r, so they determine the same sequence (a_r) and the same Young diagram.

2. Reading Jordan block sizes from the Young diagram

Suppose (for the moment) that N is already in nilpotent Jordan form with Jordan blocks of sizes

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_\ell, \qquad \sum_{i=1}^{\ell} \lambda_i = n.$$

This is a partition $\lambda \vdash n$. For a single Jordan block $J_{\lambda_i}(0)$ one checks directly that

$$\dim\ker\big(J_{\lambda_i}(0)^r\big) = \min(r, \lambda_i).$$

Therefore, for $N = \bigoplus_i J_{\lambda_i}(0)$,

$$k_r = \dim\ker(N^r) = \sum_{i=1}^{\ell} \min(r, \lambda_i). \tag{1}$$

Hence

$$a_r = k_r - k_{r-1} = \sum_{i=1}^{\ell} \Big(\min(r, \lambda_i) - \min(r-1, \lambda_i)\Big) = \#\{\, i : \lambda_i \geqslant r \,\}. \tag{2}$$

But $\#\{i : \lambda_i \geqslant r\}$ is exactly the number of boxes in the r-th *column* of the Young diagram of λ. In other words:

> The partition $\lambda'(N) = (a_1, a_2, \ldots)$ obtained from kernel growth is the *conjugate partition* λ', i.e. the Young diagram of $\lambda'(N)$ is the transpose of the Young diagram of Jordan block sizes λ.

Therefore the Jordan block sizes λ can be recovered uniquely from $\lambda'(N)$ by transposing the Young diagram.

3. Classification theorem

Theorem 1. *Let $N, M \in M_n(F)$ be nilpotent. Then N and M are similar if and only if their Jordan block size multisets coincide (equivalently, their partitions $\lambda(N)$ and $\lambda(M)$ coincide up to ordering).*

Proof. ($\Rightarrow$) If $N \sim M$, then $\dim\ker(N^r) = \dim\ker(M^r)$ for all r, so $a_r(N) = a_r(M)$ for all r. Thus $\lambda'(N) = \lambda'(M)$ as partitions. Transposing Young diagrams gives $\lambda(N) = \lambda(M)$, i.e. the Jordan block sizes coincide.

($\Leftarrow$) If N and M have the same Jordan block size partition λ, then by (1) they have the same k_r for all r, hence the same $\lambda'(N)$. But a partition λ determines a unique nilpotent Jordan matrix J_λ (a block diagonal matrix with nilpotent Jordan blocks of sizes λ_i). Since both N and M are similar to J_λ, they are similar to each other. $\square$

4. A worked example

Take $n = 7$ and suppose the nilpotent Jordan block sizes are

$$\lambda = (4, 2, 1).$$

Then (2) gives

$$a_1 = \#\{\lambda_i \geqslant 1\} = 3, \quad a_2 = \#\{\lambda_i \geqslant 2\} = 2, \quad a_3 = \#\{\lambda_i \geqslant 3\} = 1, \quad a_4 = \#\{\lambda_i \geqslant$$

and $a_r = 0$ for $r \geqslant 5$. Thus $\lambda' = (3, 2, 1, 1)$.

Compute the kernel dimensions:

$$k_1 = 3, \qquad k_2 = 3 + 2 = 5, \qquad k_3 = 6, \qquad k_4 = 7,$$

and $k_r = 7$ for $r \geqslant 4$. From the Young diagram of $\lambda' = (3, 2, 1, 1)$, transposing recovers $\lambda = (4, 2, 1)$.

One can depict the diagrams (boxes) as follows:

Young diagram of $\lambda = (4, 2, 1)$:

$$\begin{array}{llll} \square & \square & \square & \square \\ \square & \square & & \\ \square & & & \end{array}$$

Transpose (swap rows and columns) gives Young diagram of $\lambda' = (3, 2, 1, 1)$:

$$\begin{array}{lll} \square & \square & \square \\ \square & \square & \\ \square & & \\ \square & & \end{array}$$

Interpretation. The numbers $a_r = k_r - k_{r-1}$ count how many Jordan blocks have size at least r; these are precisely the column lengths of the Jordan diagram.

5.3 Relationship Between Young Diagrams and Finite Abelian p-groups

5.3.1 The Fundamental Correspondence

The relationship between **Young diagrams** (the geometric shapes of Young tableaux) and **finite abelian p-groups** is a beautiful example of structural isomorphism. Essentially, the classification of finite abelian p-groups is identical to the classification of integer partitions.

The **Fundamental Theorem of Finite Abelian Groups** states that every finite abelian p-group G is isomorphic to a direct sum of cyclic p-groups:

$$G \cong \mathbb{Z}/p^{\lambda_1}\mathbb{Z} \oplus \mathbb{Z}/p^{\lambda_2}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p^{\lambda_k}\mathbb{Z}$$

By convention, we order the exponents such that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k > 0$.

This sequence of exponents $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a **partition** of some integer n, where the total order of the group is $|G| = p^n$.

5.3.2 The Young Diagram Connection

- Each finite abelian p-group of order p^n corresponds to a unique **Young diagram** with n boxes.
- The **row lengths** of the diagram correspond to the exponents of the cyclic factors.

5.3.3 Subgroups and Conjugate Diagrams

The relationship goes deeper when you look at the **dual structure** of the group. If λ is the partition representing G, the **conjugate partition** λ' (obtained by transposing the Young diagram) also has a group-theoretic meaning.

The number of boxes in the j-th **column** of the Young diagram represents the dimension of a specific quotient space. Specifically, if we define $G[p] = \{g \in G \mid pg = 0\}$, then:

$$\dim_{\mathbb{F}_p}(p^{j-1}G/p^jG) = \lambda'_j$$

In simpler terms, the length of the j-th column tells you how many cyclic components of G have size at least p^j.

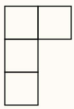
5.3.4 Summary of the Mapping

The following table summarizes how group-theoretic properties map onto the geometry of the Young diagram:

Group Property	**Young Diagram Feature**
Order of the group (p^n)	Total number of boxes (n)
Number of cyclic factors	Number of rows
Exponent of the group	Number of columns (Length of the 1st row)
Size of a specific factor (p^{λ_i})	Length of the i-th row
Rank of the group ($\dim G/pG$)	Length of the 1st column

5.3.5 Example: Order p^4

If $G \cong \mathbb{Z}/p^2\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z}$, the partition is $(2, 1, 1)$. The corresponding Young diagram is:



The **rank** of this group is 3 (the length of the first column), and its **exponent** is p^2 (the length of the first row).

6 Extensions

6.1 Problem Fa19-7: Hermite polynomials: three–term recurrence and Sturm oscillation

Let

$$\langle u, v\rangle = \int_{-\infty}^{\infty} u(x)v(x)e^{-x^2/2}\,dx$$

and let H_n be the orthogonal polynomials obtained from $1, x, x^2, \ldots$ by Gram–Schmidt. It is convenient to take the *monic* normalization: for each n let P_n be the unique monic polynomial of degree n orthogonal to all polynomials of degree $< n$. (Thus P_n is a nonzero scalar multiple of H_n, so zeros are the same.) We prove P_n has n distinct real zeros.

Three–term recurrence

Since $\deg(xP_n) = n+1$, we may expand xP_n in the orthogonal basis $\{P_0, P_1, \ldots, P_{n+1}\}$. Orthogonality forces all coefficients except those of P_{n+1}, P_n, P_{n-1} to vanish:

$$xP_n = P_{n+1} + a_n P_n + b_n P_{n-1} \qquad (n \geqslant 1),$$

hence

$$P_{n+1}(x) = (x - a_n)P_n(x) - b_n P_{n-1}(x). \tag{R}$$

Taking inner products with P_n and P_{n-1} gives

$$a_n = \frac{\langle xP_n, P_n\rangle}{\langle P_n, P_n\rangle} \in \mathbb{R}, \qquad b_n = \frac{\langle xP_n, P_{n-1}\rangle}{\langle P_{n-1}, P_{n-1}\rangle} = \frac{\langle P_n, P_n\rangle}{\langle P_{n-1}, P_{n-1}\rangle} > 0,$$

where the last equality uses $\langle xP_{n-1}, P_n\rangle = \langle P_n, P_n\rangle$ from the expansion of xP_{n-1}, and positivity uses $\langle P_m, P_m\rangle > 0$.

Thus (P_n) satisfies a three–term recurrence (R) with $b_n > 0$.

Sturm oscillation and interlacing

We prove by induction on n the stronger statement:

Claim. For each $n \geqslant 1$, P_n has n simple real zeros, and the zeros of P_{n-1} *interlace* those of P_n.

Base cases. $P_0 = 1$. Also $P_1(x) = x - a_0$ has exactly one real simple zero.

Inductive step. Assume P_n has n simple real zeros

$$\xi_1 < \xi_2 < \cdots < \xi_n$$

and the zeros of P_{n-1} interlace them. In particular, $P_{n-1}(\xi_j) \neq 0$ for all j, and the signs $P_{n-1}(\xi_j)$ alternate with j.

Evaluate the recurrence (R) at a zero ξ_j of P_n:

$$P_{n+1}(\xi_j) = (\xi_j - a_n)P_n(\xi_j) - b_n P_{n-1}(\xi_j) = -b_n P_{n-1}(\xi_j).$$

Since $b_n > 0$, the values $P_{n+1}(\xi_j)$ also alternate in sign as j increases. Therefore, by the Intermediate Value Theorem, for each $j = 1, \ldots, n-1$ there exists a point $\eta_j \in (\xi_j, \xi_{j+1})$ with

$$P_{n+1}(\eta_j) = 0.$$

So P_{n+1} has at least $n - 1$ real zeros, one in each open interval (ξ_j, ξ_{j+1}).

To get two more, use that P_{n+1} is monic, so

$$\lim_{x\to+\infty} P_{n+1}(x) = +\infty, \qquad \lim_{x\to-\infty} P_{n+1}(x) = (-1)^{n+1}\infty.$$

Since the signs $P_{n+1}(\xi_1)$ and $P_{n+1}(\xi_n)$ are fixed nonzero values, there must be one additional real zero in $(-\infty, \xi_1)$ and one in (ξ_n, ∞). Hence P_{n+1} has at least $(n-1) + 2 = n + 1$ real zeros.

But $\deg P_{n+1} = n+1$, so P_{n+1} has *exactly* $n+1$ real zeros, all simple. Moreover, we have found one zero in each interval (ξ_j, ξ_{j+1}) and one on each side, so these zeros interlace the zeros of P_n. This completes the induction.

Conclusion

Therefore P_n has n distinct real zeros for all $n \geqslant 0$. Since H_n differs from P_n only by a nonzero scalar factor, the same holds for H_n.

$$\boxed{H_n \text{ has exactly } n \text{ distinct real zeros.}}$$

7 Challenges

7.1 Proof of Uniqueness for the Extremal Function

Sp19-3: Show that $f(x) = kx(x-1)$ is the only example where equality is achieved.

Claim: The equality $\max|f(x)| = \frac{1}{8}\max|f''(x)|$ holds if and only if $f(x)$ is of the form $f(x) = Ax(1-x)$ for some constant A.

Proof. Let $M = \max_{t\in[0,1]}|f''(t)|$. Any twice continuously differentiable function f satisfying $f(0) = f(1) = 0$ can be represented using the Green's function for the operator $L = \frac{d^2}{dx^2}$:

$$f(x) = \int_0^1 G(x,t) f''(t)\, dt$$

where the Green's function is given by:

$$G(x,t) = \begin{cases} t(x-1) & 0 \leqslant t \leqslant x \\ x(t-1) & x \leqslant t \leqslant 1 \end{cases}$$

Note that $G(x,t) \leqslant 0$ for all $x, t \in [0,1]$. We can bound $|f(x)|$ as follows:

$$|f(x)| = \left|\int_0^1 G(x,t) f''(t)\, dt\right| \leqslant \int_0^1 |G(x,t)| \cdot |f''(t)|\, dt$$

Since $|f''(t)| \leqslant M$, we have:

$$|f(x)| \leqslant M \int_0^1 |G(x,t)|\, dt$$

Calculating this integral explicitly (which is equivalent to solving $y'' = -1$ with boundary conditions):

$$\int_0^1 |G(x,t)|\, dt = \frac{x(1-x)}{2}$$

Thus, the pointwise bound is:

$$|f(x)| \leqslant \frac{M}{2} x(1-x)$$

The maximum of the right-hand side occurs at $x = 1/2$, yielding the global bound $\frac{M}{8}$.

Conditions for Equality:
For the global equality $\max |f(x)| = \frac{M}{8}$ to hold, two conditions must be met:

1. **Location:** The maximum of $|f(x)|$ must occur at $x = 1/2$, because for any $x \neq 1/2$, the geometric factor $\frac{x(1-x)}{2}$ is strictly less than $1/8$.

2. **Integrand Alignment:** At $x = 1/2$, the integral inequality must be an equality:

$$\left| \int_0^1 G(1/2, t) f''(t)\, dt \right| = \int_0^1 |G(1/2, t)| \cdot M\, dt$$

   Since $G(1/2, t)$ does not change sign (it is always non-positive), for the absolute value of the integral to equal the integral of the absolute values, $f''(t)$ must have a constant sign almost everywhere. Furthermore, for $|f''(t)|$ to pull out of the integral as the constant M, we must have $|f''(t)| = M$ for all t (by continuity of f'').

Thus, $f''(t)$ must be the constant function M or $-M$. The only twice differentiable functions with constant second derivatives are quadratic polynomials. Combined with the boundary conditions $f(0) = f(1) = 0$, the solution is uniquely determined (up to the scaling factor A):

$$f(x) = Ax(1 - x)$$

□

7.2 Explicity Lagrange multipliers

Sp15-6: Find an explicit formula for the inverse of the $(n + 1) \times (n + 1)$ matrix $C(t, s)$.

7.3 Triangular decomposition

Fa20-15: Find an invertible lower triangular matrix L such that $H^{-1} = L^{\mathsf{T}} L$.

7.4 Properties implying continuity

Sp89/18 & Fa94/10 shows two condidions, that in the presence of *compactness-preserving*, implies continuity:

1. If $\{K_n\}$ is a decreasing sequence of compact subsets of $\mathbb{R}^n$, then

$$f\left(\bigcap_1^{\infty} K_n\right) = \bigcap_1^{\infty} f(K_n)$$

2. For each (x_0, y_0) in $\mathbb{R}^2$, the functions $x \mapsto f(x, y_0)$ and $y \mapsto f(x_0, y)$ are continuous.

Are the two conditions equivalent?

Answer: They are not!

(1) does not imply (2)

Let

$$g(x) = \begin{cases} \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases} \qquad f(x, y) = g(x)$$

For $a \neq 0$, $g^{-1}(a) = \{1/a\}$, and $g^{-1}(0) = \{0\}$, so each fiber of g is closed. Hence

$$f^{-1}(a) = g^{-1}(a) \times \mathbb{R}$$

is closed in $\mathbb{R}^2$. Therefore (1) holds.

However, for any y_0, the function $x \mapsto f(x, y_0) = g(x)$ is not continuous at 0. Thus (2) fails.

(2) does not imply (1)

Define

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$$

For fixed y_0, the function

$$x \mapsto \frac{x y_0}{x^2 + y_0^2}$$

is continuous, and similarly for fixed x_0. Hence (2) holds.

But along the line $y = x \neq 0$,

$$f(x, x) = \frac{x^2}{2x^2} = \frac{1}{2}$$

Thus $(0, 0)$ is a limit point of $f^{-1}(\frac{1}{2})$, yet $f(0, 0) = 0$. Therefore $f^{-1}(\frac{1}{2})$ is not closed, so (1) fails. So neither condition implies the other.

Bibliography