Jordan Canonical Form from the Standpoint of Module Theory

Yan Zeng

Version 1.0, last revised on 2014-05-07

Abstract

We give a brief summary of how to compute the Jordan canonical form from the standpoint of module theory.

Oftentimes, we need to compute the Jordan canonical form of a given square matrix A, e.g. in solving system of linear ordinary differential equations. The method given in the textbook, although conceptually appealing, is still complicated in practice. We hope to find an easier method.

丘维声 [3, page 25, 42, 68] (Reading Materials 3, 5, 7 [阅读材料三、五、七]) explained a way to compute the Jordan canonical form of A through the elementary divisors of the λ-matrix $A(\lambda) := \lambda I - A$. This is essentially the factorization theorem of a finitely generated module over the principal ideal domain $K[\lambda]$, where $K[\lambda]$ is the polynomial ring over the field K. A systematic exposition of this theory can be found in 聂灵沼和丁石孙 [2] (Chapter 5 and 6) and 龚昇 [1].

In this short notes, we will summarize the relevant results for future reference. The exposition combines materials from 龚昇 [1] and 丘维声 [3]: the theoretical aspect is taken from 龚昇 [1], and the computational aspect is taken from 丘维声 [3].

Contents

1 Introduction

Before we dive into abstract theory, we explain the prototype that motivates the theory. Let V be a finite dimensional linear vector space over the complex field $\mathbb{C}$. Let A be a linear operator on V. We want to choose a suitable set of basis for V such that the matrix representation of A has a nice form, e.g. diagonal, block diagonal, etc.

The key observation is that the matrix of A has a block diagonal form $\text{diag}\{D_1, D_2, \cdots, D_k\}$ if and only if V can be decomposed into the direct sum of subspaces invariant under A:

$$V = V_1 \bigoplus V_2 \bigoplus \cdots \bigoplus V_k.$$

Hamilton-Caley Theorem says $P(A)$ is the zero linear transformation on V (i.e. $P(A)v = 0$, $\forall v \in V$), where $P(\lambda)$ is the characteristic polynomial $|\lambda I - A|$. In other words, V is the kernel of $P(A)$. This immediately reduces the decomposition of V to the factorization of $P(\lambda)$ into irreducible elements in the polynomial ring $K[\lambda]$. Note $K[\lambda]$ is a principal ideal domain, that is, something behaves like the ring of integers. Thus, we will have at our disposal the unique factorization theorem of a principal ideal domain into prime elements.

Extending those concrete objects to an abstract setting, we come to the standpoint of representation theory: *linear vector spaces equipped with linear transformations are modules over principal ideal domains.* Therefore, the (concrete) decomposition of V into direct sum of subspaces invariant under A is generalized to the decomposition of a finitely generated module over a principal ideal domain into the direct sum of primary submodules.

With all these specific objects in mind (linear space, polynomial ring, characteristic polynomial, integers, prime numbers, etc.), we are ready to understand the module theory's standpoint.

2 Decomposition of a finitely generated module over a principal ideal domain

For the definitions of *ring*, *integral domain* and *principal ideal domain*, we refer to 龚昇 [1], Chapter 1. Roughly speaking, they are abstract generalizations of the integer ring $\mathbb{Z}$ and the polynomial ring $K[\lambda]$ over a field K. Here, the nontrivial conceptual leap is the move from integral domain to principal ideal domain. Integer ring $\mathbb{Z}$ and the polynomial ring $K[\lambda]$ are principal ideal domains because of Eulidean algorithm. The Unique Factorization Theorem for $\mathbb{Z}$ and $K[\lambda]$ can then be generalized to principal ideal domain. That's sufficient for our purpose of decomposing a linear vector space.

The connection between Unique Factorization Theorem of $K[\lambda]$ and the decomposition of V is made by the concept of module. More precisely, define the action of $f(\lambda) \in K[\lambda]$ on V by

$$f(\lambda)v = f(A)v, \ \forall v \in V.$$

If $V = \ker(f(\lambda))$ and $f(\lambda)$ has the factorization into non-associated prime elements $f(\lambda) = \prod_{i=1}^{n}[p_i(\lambda)]^{e_i}$, we have $V = \bigoplus_{i=1}^{n} \ker(p_i^{e_i})$.

For a general module, various technical conditions are needed to produce a clean-cut form of Unique Factorization Theorem. The relevant results are listed below. For all the jargons, results, and proofs, see 龚昇 [1], Chapter 4.

Theorem 2.1. *(***Cyclic decomposition theorem of a finitely generated module over a principal ideal domain − in terms of elementary divisors***) Suppose M is a non-zero finitely generated module over a principal ideal domain R. Then*

$$M = M_{tor} \bigoplus M_{free},$$

where M_{tor} is the collection of torsion elements in M and M_{free} is a free module, whose rank is uniquely determined by M.

$ann(M) = \{r \in R : rM = \{0\}\}$ is an ideal of R and hence is generated by some element $p_1^{e_1} \cdots p_n^{e_n} \in R$. Here p_i's are prime elements which are not associated to each other. Then

$$M_{tor} = M_{p_1} \bigoplus \cdots \bigoplus M_{p_n},$$

where $M_{p_i} = \{v : p_i^{e_i} v = 0\}$ is a primary module with order $p_i^{e_i}$, $i = 1, \cdots, n$.

Each M_{p_i} can be further decomposed into the sum of cyclic submodules

$$M_{p_i} = C_{i,1} \bigoplus \cdots \bigoplus C_{i,k_i},$$

where $C_{i,j}$ has order $p_i^{e_{i,j}}$ $(j = 1, \cdots, k_i)$ and

$$e_i = e_{i,1} \geq e_{i,2} \geq \cdots \geq e_{i,k_i} \geq 1, \ i = 1, \cdots, n.$$

$p_i^{e_{i,j}}$'s are called **elementary divisors** *and are uniquely determined by M up to the equivalence of associativity. In summary, we have*

$$M = (C_{1,1} \bigoplus \cdots \bigoplus C_{1,k_1}) \bigoplus \cdots \bigoplus (C_{n,1} \bigoplus \cdots \bigoplus C_{n,k_n}) \bigoplus M_{free}.$$

Note if S and T are cyclic submodules of M with $\text{ann}(S) = \langle a \rangle$ and $\text{ann}(T) = \langle b \rangle$, and $\gcd(a, b) = 1$, then $S \cap T = \{0\}$ and $S \bigoplus T$ is also a cyclic submodule with $\text{ann}(S \bigoplus T) = \langle ab \rangle$. With this observation, let $D_1 = C_{1,1} \bigoplus \cdots \bigoplus C_{n,1}$, then D_1 is a cyclic submodule with order $q_1 = \prod_{i=1}^n p_i^{e_{i,1}}$. Define D_2, $\cdots$, D_m similarly, where $m = \max_i(k_i)$. More precisely, we can arrange the elementary divisors into the following array

$$\begin{array}{ccc} p_1^{e_{1,1}} & \cdots & p_1^{e_{1,m}} \\ p_2^{e_{2,1}} & \cdots & p_2^{e_{2,m}} \\ \cdots & \cdots & \cdots \\ p_n^{e_{n,1}} & \cdots & p_n^{e_{n,m}} \end{array}$$

where $e_{i,j} = \begin{cases} e_{i,j}, & \text{if } j \leq k_i \\ 0, & \text{otherwise} \end{cases}$. Define $q_j = \prod_{i=1}^m p_i^{e_{i,j}}$, $j = 1, \cdots, m$. We also arrange the invariant subspaces into the following array

$$\begin{array}{ccc} C_{1,1} & \cdots & C_{1,m} \\ C_{2,1} & \cdots & C_{2,m} \\ \cdots & \cdots & \cdots \\ C_{n,1} & \cdots & C_{n,m} \end{array}$$

where $C_{i,j} = \begin{cases} C_{i,j}, & \text{if } j \leq k_i \\ \{0\}, & \text{otherwise} \end{cases}$. Define $D_j = \bigoplus_{i=1}^m C_{i,j}$, $j = 1, \cdots, m$.

Theorem 2.2. *(***Cyclic decomposition theorem of a finitely generated module over a principal ideal domain – in terms of invariant factors***) Suppose M is a finitely generated module over a principal ideal domain R, then*

$$M = D_1 \bigoplus \cdots \bigoplus D_m \bigoplus M_{free},$$

where M_{free} is a free submodule of M, and D_j is a cyclic submodule of M with order q_j, $j = 1, \cdots, m$. Moreover

$$q_m | q_{m-1}, q_{m-1} | q_{m-2}, \cdots, q_2 | q_1.$$

q_i, $i = 1, \cdots, m$ are called **invariant factors** *of M, which are uniquely determined by M up to the equivalence of associativity.*

3 Decomposition of a finite dimensional linear vector space under a linear operator

Now we come back to our original problem: *given a linear vector space V over a field K, for a given linear operator A on V, find a suitable set of basis for V under which the matrix representation of A has a nice form.*

In view of the module theory presented in the previous section, we note the polynomial ring $K[\lambda]$ is a principal ideal domain and V is a finitely generated torsion module over $K[\lambda]$ with the action defined by $p(\lambda)a = p(A)a$, $\forall p(\lambda) \in K[\lambda]$ and $a \in A$. $\mathrm{ann}(V) = \{p(\lambda) \in K[\lambda] : p(\lambda)V = \{0\}\}$ is an ideal of $K[\lambda]$, so there exists a unique $m(\lambda) \in K[\lambda]$ such that the coefficient of the term with highest degree is 1 and $\mathrm{ann}(V) = \langle m(\lambda)\rangle$. $m(\lambda)$ is called the **minimal polynomial** of A.

Theorem 3.1. *Suppose V is a finite dimensional linear vector space and A is a linear operator on V. Suppose the minimal polynomial $m(\lambda)$ of A has the decomposition into prime elements*

$$m(\lambda) = p_1^{e_1}(\lambda) \cdots p_n^{e_n}(\lambda),$$

where $p_i(\lambda)$, $i = 1, \cdots, n$ are non-associated monic order polynomial. Then V can be decomposed into direct sum

$$V = V_{p_1} \bigoplus \cdots \bigoplus V_{p_n},$$

where $V_{p_i} = \{v \in V : p_i^{e_i}(\lambda)v = 0\}$ and the minimal polynomial of $A|_{V_{p_i}}$ is $p_i^{e_i}(\lambda)$, $i = 1, \cdots, n$.

Each V_{p_i} $(i = 1, \cdots, n)$ can be further decomposed into the direct sum of cyclic subspace

$$V_{p_i} = \langle v_{i,1}\rangle \bigoplus \cdots \bigoplus \langle v_{i,k_i}\rangle$$

where $A|_{\langle v_{i,j}\rangle}$ has minimal polynomial $p_i^{e_{i,j}}(\lambda)$ $(j = 1, \cdots, k_i)$ and

$$e_i = e_{i,1} \geq e_{i,2} \geq \cdots \geq e_{i,k_i} \geq 1.$$

The elementary divisors $p_i^{e_{i,j}}(\lambda)$ of V are uniquely determined by A. In summary, we have

$$V = (\langle v_{1,1}\rangle \bigoplus \cdots \bigoplus \langle v_{1,k_1}\rangle) \bigoplus \cdots \bigoplus (\langle v_{n,1}\rangle \bigoplus \cdots \bigoplus \langle v_{n,k_n}\rangle).$$

If $deg p_i^{e_{i,j}}(\lambda) = d_{i,j}$, then

$$\mathcal{B}_{i,j} = (v_{i,j}, Av_{i,j}, \cdots, A^{d_{i,j}-1}v_{i,j})$$

is a basis of $\langle v_{i,j}\rangle$. The matrix of A relative to the basis $\mathcal{B} = (\mathcal{B}_{1,1}, \cdots, \mathcal{B}_{n,k_n})$ is a block diagonal matrix

$$[A]_{\mathcal{B}} = \begin{bmatrix} C[p_1^{e_{1,1}}(\lambda)] & & & & & & \\ & \ddots & & & & & \\ & & C[p_1^{e_{1,k_1}}(\lambda)] & & & & \\ & & & \ddots & & & \\ & & & & C[p_n^{e_{n,1}}(\lambda)] & & \\ & & & & & \ddots & \\ & & & & & & C[p_n^{e_{n,k_n}}(\lambda)] \end{bmatrix}$$

where

$$C[p(x)] := \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

for $p(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$. The block diagonal matrix $[A]_{\mathcal{B}}$ is called the **rational canonical form** *of A.*

As a by-product, by using the rational canonical form of A, we can see the characteristic polynomial $|\lambda I - A|$ is equal to the product of elementary divisors $\prod_{i,j} p_i^{e_{i,j}}(\lambda)$.

When K is algebraically closed, the minimal polynomial $m(\lambda)$ of A can be decomposed into product of linear factors. This allows us to get a simpler form of the matrix representation of A.

Theorem 3.2. *If the minimal polynomial $m(\lambda)$ of A can be split on K, i.e.*

$$m(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_n)^{e_n},$$

then V can be decomposed into

$$V = (\langle v_{1,1}\rangle \bigoplus \cdots \bigoplus \langle v_{1,k_1}\rangle) \bigoplus \cdots \bigoplus (\langle v_{n,1}\rangle \bigoplus \cdots \bigoplus \langle v_{n,k_n}\rangle),$$

where $\langle v_{i,j}\rangle$ is a cyclic subspace of V and the minimal polynomial of $A|_{\langle v_{i,j}\rangle}$ is $(x-\lambda_i)^{e_{i,j}}$ with

$$e_i = e_{i,1} \geq e_{i,2} \geq \cdots \geq e_{i,k_i} \geq 1.$$

These elementary divisors are uniquely determined by A. Let

$$\mathcal{G}_{i,j} = (v_{i,j}, (A-\lambda_i)v_{i,j}, \cdots, (A-\lambda_i)^{e_{i,j}-1}v_{i,j}).$$

Then $\mathcal{G}_{i,j}$ is a basis for $\langle v_{i,j}\rangle$ and the matrix representation of A under the basis $\mathcal{G} = (\mathcal{G}_{1,1}, \cdots, \mathcal{G}_{n,k_n})$ is the **Jordan canonical form**

$$[A]_{\mathcal{B}} = \begin{bmatrix} g(\lambda_1, e_{1,1}) & & & & & & \\ & \ddots & & & & & \\ & & g(\lambda_1, e_{1,k_1}) & & & & \\ & & & \ddots & & & \\ & & & & g(\lambda_n, e_{n,1}) & & \\ & & & & & \ddots & \\ & & & & & & g(\lambda_n, e_{n,k_n}) \end{bmatrix},$$

where $g(\lambda_i, e_{i,j})$ is the $e_{i,j} \times e_{i,j}$ matrix

$$g(\lambda_i, e_{i,j}) = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & & \lambda_i \end{bmatrix}.$$

So far so good, but a fundamental question remains unanswered: *how do we find the elementary divisors so that we can write out directly the Jordan canonical form?* This is the content of the next section.

4 Computation of elementary divisors and invariant factors

This section is based on 丘维声 [3] (Reading Material 3, 5 and 7). Suppose R is a ring and we can define a matrix over R, whose entries are elements of R. Elementary row operations of matrices over R include swapping two rows, multiplying a row by an invertible element of R, and adding the multiple of one row to another row. Elementary column operations are defined similarly. By Cramer's rule, a matrix over R is invertible if and only if its determinant is an invertible element of R.

In particular, we consider the case of $R = \mathbb{C}[\lambda]$, the polynomial ring over the complex number field $\mathbb{C}$. Note the collection of invertible elements of $\mathbb{C}[\lambda]$ is just $\mathbb{C}$. For any square matrix A, we would like to find its elementary divisors.

丘维声 [3] defined the invariant factors and elementary divisors for $A(\lambda) = \lambda I - A$. The invariant factors and elementary divisors thus defined are the same as those defined in 龚昇 [1]. However, proofs that these differently defined concepts are the same cannot be found in both sources. I don't find a nice exposition elsewhere of the equivalence at this moment, so I'll only focus on the algorithmic aspect which allows us to calculate things explicitly.

Theorem 4.1. *Let A be a given square matrix over the complex field $\mathbb{C}$. Define $A(\lambda) = \lambda I - A$. Through the elementary row and column manipulation of $A(\lambda)$ as an element of the ring $\mathbb{C}[\lambda]$, $A(\lambda)$ can be reduced to a diagonal matrix. Each entry on the diagonal line can be factorized into the form of $(\lambda - \lambda_1)^{p_1}(\lambda - \lambda_2)^{p_2} \cdots (\lambda - \lambda_k)^{p_k}$, and each $(\lambda - \lambda_i)^{p_i}$ is an elementary divisor of $A(\lambda)$, corresponding to the $p_i \times p_i$ Jordan block*

$$\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}.$$

5 Examples

Example 1. $A = \begin{bmatrix} 2 & 3 & 2 \\ 1 & 8 & 2 \\ -2 & -14 & -3 \end{bmatrix}$. Its λ-matrix can be reduced to

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-3)^2 & 0 \\ 0 & 0 & \lambda - 1 \end{bmatrix}.$$

So the elementary divisors are $(\lambda - 3)^2$ and $(\lambda - 1)$, and the minimal polynomial is $(\lambda - 3)^2(\lambda - 1)$. The Jordan canonical form of A is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix},$$

which can be verified by the **jordan** function of Matlab or the **JordanDecomposition** function of Mathematica. The invariant factors can be obtained by looking up the following array (see Theorem 2.2 or 丘维声 [3], Reading Material 7)

$$\begin{array}{ccc} (\lambda-3)^2 & 1 & 1 \\ (\lambda-1) & 1 & 1 \end{array}$$

So $d_1(\lambda) = d_2(\lambda) = 1$ and $d_3(\lambda) = (\lambda - 3)^2(\lambda - 1)$.

Example 2. $A = \begin{bmatrix} 3 & -4 & 0 & 2 \\ 4 & -5 & -2 & 4 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 2 & -1 \end{bmatrix}$. Its λ-matrix can be reduced to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (\lambda+1)^2 & 0 \\ 0 & 0 & 0 & (\lambda-1)^2 \end{bmatrix}.$$

So the elementary divisors are $(\lambda + 1)^2$ and $(\lambda - 1)^2$, and the minimal polynomial is $(\lambda + 1)^2(\lambda - 1)^2$. The Jordan canonical form of A is

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

The invariant factors are $d_1(\lambda) = d_2(\lambda) = d_3(\lambda) = 1$ and $d_4(\lambda) = (\lambda+1)^2(\lambda-1)^2$, obtained by looking up the following array

$$\begin{array}{cccc} (\lambda+2)^2 & 1 & 1 & 1 \\ (\lambda-1)^2 & 1 & 1 & 1 \end{array}$$

Example 3. $A = \begin{bmatrix} 3 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 3 & 0 & 5 & -3 \\ 4 & -1 & 3 & -1 \end{bmatrix}$. Its λ-matrix can be reduced to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & (\lambda-2)^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & (\lambda-2)^2 \end{bmatrix}.$$

So the elementary divisors are $(\lambda-2)^2$ and $(\lambda-2)^2$, and the minimal polynomial is $(\lambda-2)^2$. The Jordan canonical form of A is

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

The invariant factors are $d_1(\lambda) = d_2(\lambda) = 1$, $d_3(\lambda) = d_4(\lambda) = (\lambda-2)^2$, obtained by looking up the following array

$$\begin{array}{cccc} (\lambda-2)^2 & (\lambda-2)^2 & 1 & 1 \end{array}$$

References

[1] 龚昇：《线性代数五讲》，科学出版社，北京，2004。[Gong Sheng. *Five lectures in linear algebra* (in Chinese), Science Press, Beijing, 2004.] 1, 2, 6

[2] 聂灵沼、丁石孙：《代数学引论（第二版）》，高等教育出版社，北京，2000。[Nie Ling-Zhao and Ding Shi-Sun. *Introduction to algebra* (in Chinese), 2nd Edition, Higher Education Press, 2000.] 1

[3] 丘维声：《高等代数（下册）》，高等教育出版社，北京，1996。[Qiu Wei-Sheng. *Advanced algebra* (in Chinese), Volume 2, Higher Education Press, 1996.] 1, 5, 6