

A Geometric Proof for the Jordan Canonical Form of a matrix A

Xiao-Biao Lin

Department of mathematics
North Carolina State University
Raleigh, NC 27695-8205

Let A be an $n \times n$ matrix. We say that u is an eigenvector corresponding to the eigenvalue λ if

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

We say u is a generalized eigenvector if there exists $N > 1$ such that

$$(A - \lambda I)^N u = 0.$$

The smallest of such N is called the flag of the generalized eigenvector u . If u is a generalized eigenvector of flag N , then the sequence of vectors

$$u, (A - \lambda I)u, \dots, (A - \lambda I)^{N-1}u,$$

are all generalized eigenvectors of descending flags. The last entry $(A - \lambda I)^{N-1}u$ is of flag one and is just a regular eigenvector. One can verify that the vectors in the sequence are all linearly independent.

Theorem 0.1. *[Jordan Canonical form] There exists a basis of n generalized eigenvectors*

$$P = (u_1, \dots, u_n)$$

such that

$$P^{-1}AP = \text{diag}(B_i), \quad i = 1, \dots, m$$

where B_i is a square matrix of the form

$$B_i = \begin{pmatrix} \lambda_i & 1 & 0 \dots 0 \\ 0 & \lambda_i & 1 \dots 0 \\ 0 & \dots & \lambda_i \dots 1 \\ 0 & 0 & 0 \dots \lambda_i \end{pmatrix} = \text{diag}(\lambda_i) + N_i.$$

The matrices $N_i, i = 1, \dots, m$, is nilpotent.

See http://en.wikipedia.org/wiki/Nilpotent_matrix

The proof of Theorem 0.1 is divided into several lemmas, which are of independent interest.

For each $\lambda \in \mathbb{C}$, let $A_\lambda = A - \lambda I$. Let $X = \mathbb{C}^n$.

Lemma 0.2.

$$X \supset R(A_\lambda) \supset \dots \supset R(A_\lambda^k) \supset R(A_\lambda^{k+1}), \\ R(A_\lambda^{k+1}) = A_\lambda R(A_\lambda^k)$$

Lemma 0.3.

$$\begin{aligned} 0 \in N(A_\lambda) \subset N(A_\lambda^2) \subset \cdots \subset N(A_\lambda^k) \subset N(A_\lambda^{k+1}), \\ N(A_\lambda^{k+1}) = A_\lambda^{-1}N(A_\lambda^k), \end{aligned}$$

where A_λ^{-1} denotes the pre-image of the mapping A_λ .

Definition 0.4. There exists $k \geq 0$ such that $N(A_\lambda^k) = N(A_\lambda^{k+1})$. The smallest such number is called the ascent of A_λ , and is denoted by $\alpha(A_\lambda)$.

There exists $k \geq 0$ such that $R(A_\lambda^k) = R(A_\lambda^{k+1})$. The smallest such number is called the descent of A_λ , and is denoted by $\delta(A_\lambda)$.

Remark 0.5. (1) If $\alpha(A_\lambda) = 0$, then A_λ is nonsingular. If $\delta(A_\lambda) = 0$, then $R(A_\lambda) = X$. Again A_λ is nonsingular. Thus, $\alpha(A_\lambda) = 0$ is equivalent to $\delta(A_\lambda) = 0$.

(2) $R(A_\lambda^k) = R(A_\lambda^{k+1})$ for all $k \geq \delta(A_\lambda)$. Similarly $N(A_\lambda^k) = N(A_\lambda^{k+1})$ for all $k \geq \alpha(A_\lambda)$.

Lemma 0.6. $\alpha := \alpha(A_\lambda) = \delta := \delta(A_\lambda)$.

Proof. The proof is divided into two parts:

(1) Show $\alpha \leq \delta$.

Since $N(A_\lambda^{\alpha-1}) \neq N(A_\lambda^\alpha)$, there exists $x_0 \in N(A_\lambda^\alpha)$ but $x_0 \notin N(A_\lambda^{\alpha-1})$. Thus, $y = A_\lambda^{\alpha-1}x_0 \neq 0$ but $A_\lambda^\alpha x_0 = 0$. This shows that $\dim R(A_\lambda^\alpha) < \dim R(A_\lambda^{\alpha-1})$. Thus $R(A_\lambda^{\alpha-1}) \neq R(A_\lambda^\alpha)$. Therefore $\delta(A_\lambda) \geq \alpha(A_\lambda)$.

(2) Show $\delta \leq \alpha$.

Since $R(A_\lambda^{\delta-1}) \neq R(A_\lambda^\delta)$, there exists $y \in R(A_\lambda^{\delta-1})$, $y \neq 0$ but $A_\lambda y = 0$. Let $y = A_\lambda^{\delta-1}x_0$. Then $A_\lambda^{\delta-1}x_0 \neq 0$ but $A_\lambda^\delta x_0 = 0$. This shows that $N(A_\lambda^{\delta-1}) \neq N(A_\lambda^\delta)$. Therefore $\alpha(A_\lambda) \geq \delta(A_\lambda)$. \square

Lemma 0.7. Let λ be an eigenvalue of A , and $p := \alpha(A_\lambda) = \delta(A_\lambda)$. Then:

(1) The matrix A_λ is nonsingular in $R(A_\lambda^p)$.

(2) Let $\mu \neq \lambda$. Then the matrix A_μ is nonsingular in $N(A_\lambda^p)$.

Proof. (1) If there exists $x \in R(A_\lambda^p)$ such that $A_\lambda x = 0$, then $A_\lambda R(A_\lambda^p) \neq R(A_\lambda^p)$, contradicting to $p = \delta(A_\lambda)$.

(2) Assume that there exists $x \in N(A_\lambda^p)$ such that $A_\mu x = 0$. Since $x \in N(A_\lambda^p)$, there exists an integer $k \geq 0$ such that $A_\lambda^{k-1}x \neq 0$, $A_\lambda^k x = 0$. Applying A_λ^{k-1} to

$$A_\lambda x + (\lambda - \mu)x = 0,$$

we have $(\lambda - \mu)A_\lambda^{k-1}x = 0$. This can occur only if $\lambda - \mu = 0$. \square

Our next main result shows that X is a direct sum of two invariant subspaces $R(A_\lambda^p) \oplus N(A_\lambda^p) = X$, where λ is nonsingular on $R(A_\lambda^p)$ and λ is the only eigenvalue for $N(A_\lambda^p)$.

Lemma 0.8. *Let $p = \alpha(A_\lambda) = \delta(A_\lambda)$. Then*

$$R(A_\lambda^p) \oplus N(A_\lambda^p) = X.$$

Proof. (1) We show that $R(A_\lambda^p) \cap N(A_\lambda^p) = 0$.

If $y \in R(A_\lambda^p) \cap N(A_\lambda^p)$, then $y = A_\lambda^p x$ and $A_\lambda^p y = 0$. Thus $A_\lambda^{2p} x = 0$. $x \in N(A_\lambda^{2p}) = N(A_\lambda^p)$. Thus $A_\lambda^p x = 0$. This implies $y = 0$.

(2) We show that for any $x \in X$, there exist $x_1 \in R(A_\lambda^p)$, $x_2 \in N(A_\lambda^p)$ such that $x = x_1 + x_2$.

Consider $A_\lambda^p x \in R(A_\lambda^p) = R(A_\lambda^{2p})$. There exists y such that $A_\lambda^{2p} y = A_\lambda^p x$. Thus

$$A_\lambda^p(x - A_\lambda^p y) = 0.$$

Let $x_1 = A_\lambda^p y$, and $x_2 = x - x_1$. Then $x_1 \in R(A_\lambda^p)$, $x_2 \in N(A_\lambda^p)$. \square

Let

$$\lambda_1, \dots, \lambda_m$$

be a list of all the distinct eigenvalues of A , each associated with a unique

$$p_i = \alpha(A_{\lambda_i}) = \delta(A_{\lambda_i}).$$

We have shown

$$R(A_{\lambda_i}^{p_i}) \oplus N(A_{\lambda_i}^{p_i}) = X.$$

Clearly $N(A_{\lambda_i}^{p_i})$ is the generalized eigenspace for λ_i .

We now show

Lemma 0.9.

$$\bigoplus_i N(A_{\lambda_i}^{p_i}) = X.$$

Proof. (1) We show that the generalized eigenspaces, each corresponding to a distinct λ_i , are linearly independent. Let k be the smallest integer such that $x_1 + \dots + x_k = 0$ where we assume, without loss of generality, $x_j \in N(A_{\lambda_j}^{p_j})$ and is nonzero.

Applying $A_{\lambda_1}^{p_1}$ to $x_1 + \dots + x_k = 0$, we have

$$A_{\lambda_1}^{p_1} x_2 + \dots + A_{\lambda_1}^{p_1} x_k = 0.$$

From Lemma 0.7, $A_{\lambda_1}^{p_1}$ is nonsingular in each of its invariant subspace $N(A_{\lambda_j}^{p_j})$, $j \neq 1$, and thus $y_j := A_{\lambda_1}^{p_1} x_j \neq 0$. We have $y_2 + \dots + y_k = 0$. This is a contradiction to the minimal property of k .

(2) For any $x \in X$, we show, by induction, that it is possible to express x as $x = x_1 + \dots + x_m$ with $x_i \in N(A_{\lambda_i}^{p_i})$.

If $m = 1$, then $R(A_{\lambda_1}^{p_1}) \oplus N(A_{\lambda_1}^{p_1})$. From Lemma 0.7, λ_1 is not an eigenvalue of A restricted to $R(A_{\lambda_1}^{p_1})$. If $R(A_{\lambda_1}^{p_1}) \neq 0$, then the restriction of A to it has no eigenvalue. This is a contradiction. Thus $R(A_{\lambda_1}^{p_1}) = 0$ and $X = N(A_{\lambda_1}^{p_1})$.

If the lemma is true for $m - 1$, then consider all the invariant subspaces

$$N(A_{\lambda_2}^{p_2}), \dots, N(A_{\lambda_m}^{p_m}).$$

They are all contained in $R(A_{\lambda_1}^{p_1})$ because from Lemma 0.7, A_{λ_1} is nonsingular in each $N(A_{\lambda_j}^{p_j}), j \geq 2$. Therefore, if A is restricted to $R(A_{\lambda_1}^{p_1})$, it has all the eigenvalues $\lambda_2, \dots, \lambda_m$ but not λ_1 . This is due to Lemma 0.7 that λ_1 is not an eigenvalue in $R(A_{\lambda_1}^{p_1})$. Based on the induction assumption, for the case of $m - 1$ eigenvalues, we have

$$R(A_{\lambda_1}^{p_1}) = \bigoplus_{2 \leq j \leq m} N(A_{\lambda_j}^{p_j}).$$

Thus

$$X = \bigoplus_{1 \leq j \leq m} N(A_{\lambda_j}^{p_j}).$$

□

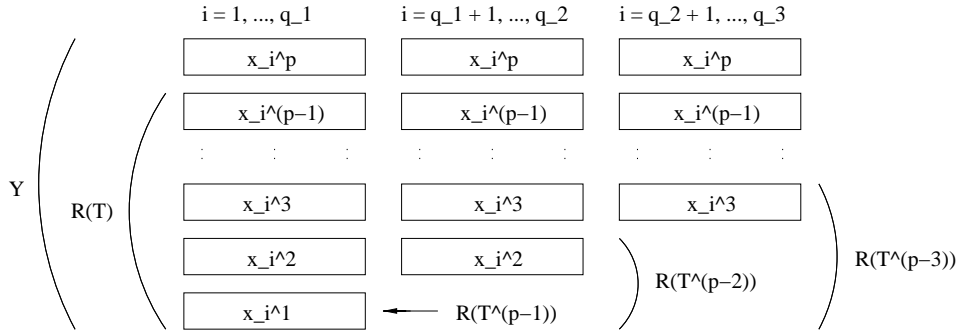


FIGURE 1. A cyclic basis can be selected for $N(A_{\lambda}^p)$.

To finish the proof of Theorem 0.1, it remains to show that in each generalized eigenspace $N(A_{\lambda_j}^{p_j})$, a cyclic basis can be selected. The idea of the proof is illustrated in Figure 1. The top, left box holds generalized eigenvectors of flag p , pick $u = x_i^p, i = 1, \dots, q_1$ from that box, then $(A_{\lambda_1} u, A_{\lambda_1}^2 u, \dots, A_{\lambda_1}^{p-1} u)$ are placed sequentially below on the the first column of boxes. Similarly the top, second-from-left box contains generalized eigenvectors of flag $p - 1$. Pick $u = x_i^p, i = q_1 + 1, \dots, q_2$ from that box and apply powers of A_{λ_2} to u generates the second column of boxes, etc. Note the second column has only $p - 1$ boxes because the flag number of the vectors from the top box is smaller.

