# (Almost) Jordan Form

These notes will demonstrate most of the basic steps for getting to **Jordan canonical form** of a complex matrix. They will also get to to an important **diagonal-nilpotent decomposition**, which we will require later. [These notes owe a tremedous debt to the beautiful notes of Ed Nelson (Princeton) which are posted on a website of Andre Reznikov (Bar-Ilan). It is worth your while to do a little internet sleuthing to find these.]

The first result is not sexy, but actually does all of the hard work.

**Theorem.** *Let $A \in \mathrm{M}_n(\mathbb{F})$ ($\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$) and $\lambda$ be an eigenvalue of $A$ in $\mathbb{F}$.*

**(i)** *There is a positive integer $m$ (geometric multiplicity) for which $\ker(A-\lambda I)^k \subseteq \ker(A-\lambda I)^m$ for each positive integer $k$. Each of the subspaces $\ker(A-\lambda I)^m$ and $\mathrm{ran}(A-\lambda I)^m$ are invariant for $A$, hence for any polynomial, $p(A)$, in $A$.*

**(ii)** *We have $\mathbb{F}^n = \ker(A-\lambda I)^m \oplus \mathrm{ran}(A-\lambda I)^m$. Hence there is $g$ in $\mathrm{GL}_n(\mathbb{F})$ for which*

$$A = g \begin{bmatrix} \lambda I_d + N & 0 \\ 0 & R \end{bmatrix} g^{-1}$$

*where $d = \dim \ker(A-\lambda I)^m$ and $N \in \mathrm{M}_d(\mathbb{F})$ with $N^m = 0$.*

**Proof. (i)** To begin with, we simply observe that

$$\ker(A-\lambda I) \subseteq \ker(A-\lambda I)^2 \subseteq \dots .$$

By finite dimensionality of $\mathbb{F}^n$, this non-decreasing chain of subspaces must stabilise; let $r$ be the smallest value for which $\ker(A-\lambda I)^m = \ker(A-\lambda I)^k$ for each $k \geq m$.

We observe that if $x \in \ker(A-\lambda I)^m$, then

$$(A-\lambda I)^m Ax = A(A-\lambda I)^m x = 0$$

so $A[\ker(A-\lambda I)^m] \subseteq \ker(A-\lambda I)^m$. Finally if $x \in \mathrm{ran}(A-\lambda I)^m$, then $x = (A-\lambda I)^m y$ for some $y$. Hence

$$Ax = A(A-\lambda I)^m y = (A-\lambda I)^m Ay \in \mathrm{ran}(A-\lambda I)^m$$

so $A[\mathrm{ran}(A-\lambda I)^m] \subseteq \mathrm{ran}(A-\lambda I)^m$. The same arguemet holds for $p(A)$.

**(ii)** If $x \in \ker(A-\lambda I)^m \cap \mathrm{ran}(A-\lambda I)^m$, then on one hand $0 = (A-\lambda I)^m x$, while on the other, $x = (A-\lambda I)^m y$ for some vector $y$. Thus $0 = (A-\lambda I)^m x =$

$(A-\lambda I)^{2m}y$, so $y \in \ker(A-\lambda I)^{2m} = \ker(A-\lambda I)^m$, whence $x = (A-\lambda I)^m y = 0$. By rank-nullity theorem, we find that $n = \dim\ker(A-\lambda I)^m + \dim\operatorname{ran}(A-\lambda I)^m$, so we find that $\mathbb{F}^n$ is a direct sum of these subspaces.

Let $B_1 = \{\xi_1, \ldots, \xi_d\}$ be a basis for $\ker(A-\lambda I)^m$, and $B_2 = \{\xi_{d+1}, \ldots, \xi_n\}$ a basis for $\operatorname{ran}(A-\lambda I)^m$. Then the restricted operator $(A-\lambda I)|_{\ker(A-\lambda I)^m}$ is nilpotent and admits matrix with respect to $B_1$ of the form $N$, with $N^m = 0$. Let $R$ be the matrix of $A|_{\operatorname{ran}(A-\lambda I)^m}$. Then if $g$ is the change of basis matrix from $B_1 \cup B_2 = \{\xi_1, \ldots, \xi_n\}$ to the standard basis, we get the desired result. $\square$

The following is essentially a simple induction on the "remainder" block $R$ from the theorem above. The details are left to the reader.

We take it for granted that a complex matrix admits at least one eigenvalue and at least one complex eigenvector.

**Corollary (Almost Jordan Decoposition).** *Let $A \in \mathrm{M}_n(\mathbb{C})$ and $\lambda_1, \ldots, \lambda_s$ be a full list of distinct eigenvalues for $A$ ($s$ is the size of the spectrum). Let $m_i$ be so $\ker(A-\lambda_i I)^{m_i} \supseteq \ker(A-\lambda_i I)^k$ for any positive integer $k$, and $d_i = \dim\ker(A-\lambda_i I)^{m_i}$. Then there are nilpotent matrices $N_i$ in $\mathrm{M}_{d_i}(\mathbb{C})$ with $N_i^{m_i} = 0$ and a $g$ in $\mathrm{GL}_n(\mathbb{C})$ for which*

$$A = g \begin{bmatrix} \lambda_1 I_{d_1} + N_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 I_{d_1} + N_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_s I_s + N_s \end{bmatrix} g^{-1}. \qquad (\heartsuit)$$

*Furthermore, if all eigenvalues are in $\mathbb{R}$, then we can arrange that $g \in \mathrm{GL}_n(\mathbb{R})$, as well.*

If one is willing to invest the extra effort to show that a $d \times d$ nilpotent matrix $N$ is similar to one of the form

$$\begin{bmatrix} 0 & \eta_1 & 0 & \ldots & 0 \\ 0 & 0 & \eta_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & \eta_{d-1} \\ 0 & \ldots & 0 & 0 & 0 \end{bmatrix}$$

where $\eta_1, \ldots \eta_{d-1} \in \{0, 1\}$ then she has effectivley shown the usual Jordan form. In fact if $m$ is the smallest integer for which $N^m = 0$, then there are

$\eta_i, \eta_{i+1}, \ldots, \eta_{i+m}$ which are all 1, and no consecutive chain of such $\eta_i = 1$ may be longer than $n$.

Observe that if a matrix admits the form of a block decomposition

$$A = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & A_s \end{bmatrix} \qquad (\diamondsuit)$$

then for any polynomial $p(z)$ we have

$$p(A) = \begin{bmatrix} p(A_1) & 0 & \ldots & 0 \\ 0 & p(A_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & p(A_s) \end{bmatrix}.$$

**Corollary (Almost Cayley-Hamilton Theorem).** *Given $A$ in $\mathrm{M}_n(\mathbb{C})$, as above, the polynomial $\mu_A(z) = \prod_{k=1}^{s}(z - \lambda_i)^{m_i}$ satisfies $\mu_A(A) = 0$.*

**Proof.** Following ($\heartsuit$) and then ($\diamondsuit$), we see that

$$\mu_A(A) = g\,\mu_A\left(\begin{bmatrix} \lambda_1 I_{d_1} + N_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 I_{d_1} + N_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_s I_s + N_s \end{bmatrix}\right) g^{-1}$$

$$= g \begin{bmatrix} \mu_A(\lambda_1 I_{d_1} + N_1) & 0 & \ldots & 0 \\ 0 & \mu_A(\lambda_2 I_{d_1} + N_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \mu(\lambda_s I_{d_s} + N_s) \end{bmatrix} g^{-1}.$$

Each block contains a factor $[(\lambda_k I_{d_k} + N_k) - \lambda_k I_{d_k}]^{m_k} = {N_k}^{m_k} = 0$ and is thus 0. □

The polynomial $\mu_A$ is the minimal polynomial of $A$.

We obtain a factorisation of the block decomposition given in ($\diamondsuit$)

$$A = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & I_{d_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & I_{d_s} \end{bmatrix} \begin{bmatrix} I_{d_1} & 0 & \ldots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & I_{d_s} \end{bmatrix} \cdots \begin{bmatrix} I_{d_1} & 0 & \ldots & 0 \\ 0 & I_{d_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & A_s \end{bmatrix}$$

from which it easy follows that $\det A = \prod_{k=1}^{s} \det A_s$. Hence if one is willing to show that $\det(\lambda I_d + N) = \lambda^d$, whenever $N$ is a $d \times d$ nilpotent matrix (this will follow form a form of Engel's Theorem, later in the course), then it is an easy step to show that the characteritic polynomial of $A$, above, is $p_A(z) = \prod_{k=1}^{s}(z - \lambda_i)^{d_i}$ and hence the Cayley-Hamilton Theorem holds.

Now for a different perspective on this result. We say that $A$ in $\mathrm{M}_n(\mathbb{C})$ is *diagonalisable* if there is $g$ in $\mathrm{GL}_n(\mathbb{C})$ such that

$$A = g\,\mathrm{diag}(\alpha_1, \ldots, \alpha_n)\,g^{-1}$$

for some $\alpha_1, \ldots, \alpha_n$ in $\mathbb{C}$. Notice that $A$ is diagonalisable if and only if the minimal polynomial $\mu_A(z)$, above, has multiplicity $m_k = 1$ for each $k$.

**Diagonal-Nilpotent Decomposition Theorem.** *Let $A \in \mathrm{M}_n(\mathbb{C})$. Then there is a unique decomposition*

$$A = A_D + A_N$$

*where $A_D$ is diagonalisable, $A_N$ is nilpotent, and $[A_D, A_N] = 0$. Furthermore, there are polynomials $p_D(z)$ and $p_N(z)$ for which*

$$A_D = p_D(A) \text{ and } p_N(A) = A_N.$$

**Proof.** Let us exhibit, first, such a decomposition. In the notation of ($\heartsuit$) let

$$A_D = g\begin{bmatrix} \lambda_1 I_{d_1} & 0 & \ldots & 0 \\ 0 & \lambda_2 I_{d_1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_s I_{d_s} \end{bmatrix} g^{-1} \text{ and } A_N = g\begin{bmatrix} N_1 & 0 & \ldots & 0 \\ 0 & N_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & N_s \end{bmatrix} g^{-1}.$$

Now define for each $k = 1, \ldots, s$, polynomials

$$\tilde{q}_k(z) = \prod_{\substack{j=1,\ldots,s \\ j \neq k}} (z - \lambda_j) \text{ and } q_k(z) = \frac{1}{\tilde{q}_k(\lambda_k)}\tilde{q}_k(z).$$

As in the proof of the Almost Cayley-Hamilton Theorem we compute

$$q_k(A) = g \begin{bmatrix} 0 & 0 & & \dots & & 0 & 0 \\ 0 & \ddots & & \ddots & & \ddots & 0 \\ \vdots & \ddots & & q_k(\lambda_k I_{d_k} + N_k) & & \ddots & \vdots \\ 0 & \ddots & & \ddots & & \ddots & 0 \\ 0 & 0 & & \dots & & 0 & 0 \end{bmatrix} g^{-1}$$

Observe that $q_k(\lambda_k + z)$ is simply a polynomial with constant constant coefficient 1, and hence $q_k(\lambda_k I_{d_k} + N_k) = I + r_k(N_n)$, where $r$ is a polynomial with constant coefficient 0, Hence $r_k(N_k)$ is itself, nilpotent; in fact $r_k(N_k)^{m_k} = 0$. Thus we have that

$$[I + r_k(N_k)][I - r_k(N_k) + \cdots + (-1)^{m_k-1} r_k(N_k)^{m_k-1}] = I$$

Noting that $r_k(z) = q(\lambda_k + z) - 1$ we find that the polynomial

$$p_k(z) = q(\lambda_k + z)[1 - (q(\lambda_k + z) - 1) + \cdots + (-1)^{m_k-1}(q(\lambda_k + z) - 1)^{m_k-1}]$$

satisfies $p_k(A) = gP_kg^{-1}$, where $P_k$ is block-diagonal with $I_{d_k}$ in the $k$th block and zeros elsewhere. Finally set

$$p_D(z) = \sum_{k=1}^{s} \lambda_k p_k(z)$$

and we find that $p_D(A) = A_D$. Hence $p_N(z) = p_D(z) - z$.

Now we prove uniqueness. Suppose $A = D+N$ where $D$ is diagonalisable, $N$ is nilpotent and $[D, N] = 0$. Then $[D, A] = [D, D+N] = 0$, and, similarly, $[N, A] = 0$. Thus $[D, A_D] = [D, p_D(A)] = 0$, and, similarly, $[N, A_N] = 0$. Hence the equation $A_D + A_N = A = D + N$ implies $A_D - D = N - A_N$. But then the binomial theorem implies that $(N - A_N)^{2n} = 0$, so $A_D - D = N - A_N$ is nilpotent.

Now let $E_k = gP_kg^{-1} = p_k(A)$, from above. Then $E_k^2 = E_k$ and $[D, E_k] = 0$, so $[g^{-1}Dg, P_k] = 0$. It follows, just as in the proof of the first theorem, that we have block-diagonal form

$$g^{-1}Dg = \begin{bmatrix} D_1 & 0 & \dots & 0 \\ 0 & D_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & D_s \end{bmatrix}.$$

But since the minimal polynomial $\mu_D(z)$ has multiplicites $m_i = 1$ by diagonalisability of $D$, and $\mu_D(D_k) = 0$ for eack $k$, it follows that each block $D_k$ is diagonalisable. Hence there is a block-diagonal $h$ in $\mathrm{GL}_n(\mathbb{C})$ for which $h^{-1}g^{-1}Dgh$ is diagonal. Notice that $h^{-1}g^{-1}A_D gh = g^{-1}A_D g$ remains diagonal. Hence $A_D - D$ is diagonalisable.

Thus $A_D - D$ is both nilpotent and diagonalisable, so $A_D - D = 0$. □

We say two diagonalisable $n \times n$ matrices $A$ and $B$ are *simultaneously diagonalisable* if there are complex numbers $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ and a $g$ in $\mathrm{GL}_n(\mathbb{C})$ such

$$A = g\,\mathrm{diag}(\alpha_1, \ldots, \alpha_n)\,g^{-1} \text{ and } B = g\,\mathrm{diag}(\beta_1, \ldots, \beta_n)\,g^{-1}.$$

In the course of proving the above result we showed the non-trivial direction of the following.

**Corollary.** *Two diagonalisable matrices are simultaneously diagonalisable if and only if they commute.*

Written by Nico Spronk, for use by students of PMath 763 at University of Waterloo.