# Minimal Polynomials and Jordan Normal Forms

## 1. Minimal Polynomials

Let $A$ be an $n \times n$ real matrix.

**M1.** There is a polynomial $p$ such that $p(A) = 0$.

**Proof.** The space $M_{n\times n}(\mathbb{R})$ of $n \times n$ real matrices is an $n^2$-dimensional vector space over $\mathbb{R}$. Take $I, A, A^2, \ldots, A^{n^2}$. There are $n^2 + 1$ elements here, so they are linearly dependent: $\mu_0 I + \mu_1 A + \ldots + \mu_{n^2} A^{n^2} = 0$ for some $\mu_i$, not all 0. Let $p(t) = \sum \mu_i t^i$, then $p(A) = 0$.

So we have a polynomial giving 0 for $A$. Let $m$ be a polynomial such that $m(A) = 0$ which is *monic* (leading coefficient = 1) and of *smallest degree*. This is the *minimal polynomial* of $A$.

**M2.** The minimal polynomial of $A$ is unique.

**Proof.** If we had two such polynomials, they must both have the same degree and the same leading coefficient 1, and so their difference is a polynomial of smaller degree which still gives 0 when applied to $A$. But this would contradict the minimality of $m$.

**M3.** If $p$ is some polynomial such that $p(A) = 0$, then $m$ divides $p$.

**Proof.** Clearly $\deg(p) \geqslant \deg(m)$. By polynomial division, we may write $p = qm + r$ for some polynomials $q, r$, such that $\deg(r) < \deg(m)$. Then $r(A) = p(A) - q(A)m(A) = 0$, which contradicts the minimality of $m$, unless $r = 0$. So $r = 0$, and $m$ divides $p$.

**M4.** If $\lambda$ is an eigenvalue of $A$, then it is a root of $m$.

**Proof.** We have $Av = \lambda v$, for some $v \neq 0$. Then $A^2 v = \lambda^2 v$, and continuing, $A^k v = \lambda^k v$ for any integer $k \geqslant 0$. Therefore, for any polynomial $p$, we have $p(A)v = p(\lambda)v$. In particular, we have $m(A)v = m(\lambda)v$. But $m(A) = 0$ and $v \neq 0$, thus $m(\lambda) = 0$.

**Cayley-Hamilton.** Let $\chi(t) = \det(A - tI)$ be the characteristic polynomial of $A$. Then $\chi(A) = 0$.

**Proof.** See the notes.

If we are given $\chi(t)$, we have some idea what $m(t)$ might be.

**M5.** If $\chi(t) = \prod_{i=1}^{r}(t - \lambda_i)^{a_i}$, what are the possibilities for $m(t)$?

By Cayley-Hamilton and **M3**, we know that $m$ divides $\chi$. And by **M4** we know that every root of $\chi$ must appear in $m$. Thus the only possibilities for $m$ are $\prod_{i=1}^{r}(t - \lambda_i)^{c_i}$, where $1 \leqslant c_i \leqslant a_i$.

Lastly, a useful property:

**M6.** $\chi(t)$ and $m(t)$ are invariant under change of basis.

**Proof.** Let $B = P^{-1}AP$. Since $B^k = (P^{-1}AP)^k = P^{-1}A^kP$, if $p(t)$ is some polynomial then $p(B) = P^{-1}p(A)P$. So $p(B) = 0$ iff $p(A) = 0$. Thus the minimal polynomials of $A$ and $B$ divide each other, so are equal.

Also, $\det(B - tI) = \det(P^{-1}(A - tI)P) = \det(P^{-1})\det(A - tI)\det(P) = \det(A - tI)$. so $\chi_B(t) = \chi_A(t)$.

## 2. Jordan Blocks & Jordan Normal Form

What do the characteristic and minimal polynomials tell us about the Jordan Normal Form? (Be warned that most of the following is explanation rather than proof.)

First, a *Jordan block*. These are of the form

$$\begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

for some $\lambda$, with all other entries 0.

The *Jordan Normal Form* of a matrix $A$ looks like

$$\begin{pmatrix} J & & \\ & \ddots & \\ & & J \end{pmatrix}$$

where the $J$s are Jordan block matrices of some (possibly different) sizes and with (possibly different) $\lambda$s. The actual details of what sizes and $\lambda$s we have depend upon the matrix $A$. (We are allowed to write the blocks in a different order and still call it 'the' JNF. That's a cosmetic change.)

Note: I assume here that we're working over an algebraically closed field such as $\mathbb{C}$, so that the eigenvalues and JNF actually exist.

**J1.** Let $J$ be a Jordan block with diagonal entries $\lambda$. Then $\lambda$ is the only eigenvalue, and the associated eigenspace is only 1-dimensional.

> **Proof.** Since $J$ is upper-triangular, it is clear that the only eigenvalue is $\lambda$. Solving $J\mathbf{x} = \lambda\mathbf{x}$ gives us the equations $\lambda x_i + x_{i+1} = \lambda x_i$ for $i < n$, and $\lambda x_n = \lambda x_n$, from which we see that $x_2 = \ldots = x_n = 0$, giving the eigenvector $(1, 0, \ldots, 0)$. (Or just note that $n(J - \lambda I) = 1$ since $r(J - \lambda I) = n - 1$, as the last $n - 1$ columns of $J - \lambda I$ are clearly linearly independent.)

**J2.** Let $J$ be a $k \times k$ Jordan block. Then $(J - \lambda I)^n = 0$ for $n = k$, but is non-zero for $n < k$.

> **Sketch.** If we square the matrix $(J - \lambda I)$, we get
>
> $$\begin{pmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix}$$
>
> Each subsequent multiplication by $(J - \lambda I)$ pushes the diagonal of 1s further towards the top-right corner. After $k - 1$ steps, there is a single 1 in the top-right, and after $k$ steps we have 0.

**J3.** Suppose that $A, B$ are block matrices, with the same block sizes. Then

$$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_n \end{pmatrix} \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_n \end{pmatrix} = \begin{pmatrix} A_1B_1 & & \\ & \ddots & \\ & & A_nB_n \end{pmatrix}$$

**Sketch.** Just try it. For a row/column passing through a particular block, the rest of the row/column outside the block is entirely full of 0s, so the blocks never interact when multiplying, and every entry in $AB$ outside the blocks is 0.

**J4.** Suppose a matrix $A$ is in Jordan Normal Form, containing some number of Jordan blocks down the diagonal. Suppose that the distinct diagonal entries of $A$ are $\lambda_1, \ldots, \lambda_k$, with $\lambda_i$ appearing with multiplicity $a_i$. Suppose that the number of Jordan blocks whose diagonal entries are $\lambda_i$ is $g_i$. And suppose that the largest Jordan block whose diagonal entries are $\lambda_i$ is size $c_i$.

Then we have the following results:

(i) $\chi(t) = \prod_{i=1}^{k}(t - \lambda_i)^{a_i}$

(ii) $m(t) = \prod_{i=1}^{k}(t - \lambda_i)^{c_i}$ ("the power of $(t - \lambda_i)$ in $m(t)$ is the size of the largest $\lambda_i$-block")

(iii) $n(A - \lambda_i I) = g_i$ ("the dimension of the $\lambda_i$-eigenspace is the number of $\lambda_i$-blocks")

**Sketch.**

(i) Since $A$ is in Jordan Normal Form, it is upper-triangular, and so its eigenvalues are the diagonal entries.

(ii) By the previous point and **M4** we know that $m(t) = \prod_{i=1}^{k}(t - \lambda_i)^{k_i}$ for some $k_i$.
First consider $(A - \lambda_i I)^{k_i}$ for one particular $i$. By **J3**, this is the same as replacing each Jordan block $J$ in $A$ by $(J - \lambda_i I)^{k_i}$. By **J2**, we need $k_i = c_i$ to kill off the largest $\lambda_i$-block, by which point any smaller $\lambda_i$ block has already become 0. So in $(A - \lambda_i I)^{k_i}$, the $\lambda_i$-blocks are now all 0.
Similarly, $(A - \lambda_j I)^{c_j}$ makes the $\lambda_j$-blocks all 0. So, by **J3**, $(A - \lambda_i I)^{k_i}(A - \lambda_j I)^{c_j}$ has the $\lambda_i$- and $\lambda_j$-blocks all 0. Repeating this for each $\lambda$ leaves us with the 0 matrix. At each stage, we needed $k_i$ to be as big as $c_i$ to kill off the largest block.

(iii) By **J1**, we get one dimension of eigenspace for each Jordan block, thus $n(A - \lambda_i I) = g_i$

## 3. Small Examples

**Exercise.** Let $A$ be the matrix

$$\begin{pmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & & & \\ & & & \lambda & & \\ & & & & \mu & 1 \\ & & & & & \mu \end{pmatrix}$$

Find $(A - \lambda I)^2, (A - \lambda I)^3, (A - \mu I)^2$ and show $(A - \lambda I)^3(A - \mu I)^2 = 0$. Solve $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{x} = \mu\mathbf{x}$.

**Example.** Suppose that we are told that $A$ is a $3 \times 3$ matrix, and that we are given its characteristic and minimal polynomials. Then we know the Jordan Normal Form of $A$.

If $\chi(t) = (t - \lambda)(t - \mu)(t - \nu)$, then by **M6** we know $m(t) = \chi(t)$, so by **J4**(ii) we know that the JNF is just diagonal. If $\chi(t) = (t - \lambda)(t - \mu)^2$, then by **M6** we know $m(t)$ is either $(t - \lambda)(t - \mu)$ or equals $\chi$. Then we know whether there is one $2 \times 2$ $\mu$-block or two $1 \times 1$ blocks. And if $\chi(t) = (t - \lambda)^3$, we know $m(t) = (t - \lambda)^k$ where $k = 1, 2, 3$. Again, knowing the size of the largest block determines the JNF exactly.

**Example.** Knowing $\chi$ and $m$ isn't enough if $A$ is a $4 \times 4$ matrix.

We can assume that the eigenvalues are all the same, otherwise we look at each block in turn and are done by the previous example. So $\chi(t) = (x - \lambda)^4$ for some $\lambda$. If the largest Jordan block is size 1, 3 or 4 then the JNF is forced. But if the largest block is size 2, we can write the rest of the matrix as either another block of size 2, or two blocks of size 1.

$$\begin{pmatrix} \lambda & 1 & & \\ & \lambda & & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix} \quad \begin{pmatrix} \lambda & 1 & & \\ & \lambda & & \\ & & \lambda & \\ & & & \lambda \end{pmatrix}$$

Note that those two $4 \times 4$ JNFs could have been distinguished if we had known the eigenspace dimension, since they have a different number of blocks.

**Exercise.** Suppose we are told that $A$ is a $6 \times 6$ matrix, and that we are given its characteristic and minimal polynomials, and also the dimensions of the eigenspaces. Show that the JNF is uniquely determined. Give two $7 \times 7$ matrices which have the same characteristic and minimal polynomials, and the same eigenspace dimensions, but which have distinct JNFs.

## 4. Finding the basis for a JNF

So we know what JNFs look like and some of their properties. But how do we find the change of basis that gives us them? Here's an example. Let $A$ be the matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}$$

We find that $\chi(t) = (t-2)^3$. Clearly $A - 2I \neq 0$, so $m(t)$ is either $(t-2)^2$ or $(t-2)^3$. So let's try it:

$$(A - 2I)^2 = \begin{pmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ -2 & 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So $m(t) = (t-2)^2$, which tells us that the JNF of $A$ is

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

If we are in the basis $\{e_1, e_2, e_3\}$ wrt which $A$ is in JNF, then the matrix tells us: $Ae_1 = 2e_1,\ Ae_2 = 2e_2 + e_1,\ Ae_3 = 2e_3$. Let's rewrite those equations as: $(A-2I)e_1 = 0,\ (A-2I)e_2 = e_1,\ (A-2I)e_3 = 0$. So we want two vectors that are in $\ker(A - 2I)$, and a vector that gets sent to one of those. So let's find $e_2$ first and then define $e_1$ to be $(A-2I)e_2$. That makes sure that our $e_1$ and $e_2$ behave appropriately.

So, $\ker(A - 2I) = \{2x = y\}$, by the matrix for $(A-2I)^2$ above. Let $e_2$ be some vector not in there, say $e_2 = (1, 0, 0)$. Then let $e_1 = (A - 2I)e_2 = (-2, -4, -2)$. For $e_3$ we can then just pick any linearly independent vector in $\ker(A - 2I)$, such as $e_3 = (1, 2, 3)$. Let $P$ be the change of basis matrix, then

$$P^{-1}AP = \frac{1}{8}\begin{pmatrix} 0 & -3 & 2 \\ 8 & -4 & 0 \\ 0 & -2 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix} \begin{pmatrix} -2 & 1 & 1 \\ -4 & 0 & 2 \\ -2 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

If we'd wanted a bigger block, say an $n \times n$ block for the eigenvalue $\lambda$, then we could pick any non-zero $e_n$ that isn't in $\ker(A - \lambda I)^{n-1}$. Then we can let $e_{n-1} = (A - \lambda I)e_n,\ e_{n-2} = (A - \lambda I)e_{n-1}, \ldots,\ e_1 = (A - \lambda I)e_2$. Note that $(A - \lambda I)e_1 = 0$, so $e_1$ is our single eigenvector that we expected by **J1**. These $e_i$ then behave precisely as required for the matrix to become the Jordan block.

## 5. Uniqueness

For a given matrix $A$, the JNF is unique. (Well, up to reordering its Jordan blocks, but that's not an exciting change.) Let's do an example.

Let $A$ be a matrix whose JNF is the following matrix $J$ (assuming only one eigenvalue for now):

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

As before, the number of eigenvectors gives the number of blocks. And indeed, $J-\lambda I$ has five eigenvectors. This can be seen by looking at the nullity of $J - \lambda I$, and since

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

we can see clearly that the nullity is 5, from those five columns full of 0s. Now square this, using our block behaviour from section 2:

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

This has nine columns of 0s, and so the nullity of $(J - \lambda I)^2$ equals 9. But more importantly, notice where those new columns came from. The first four blocks each provided an extra column of 0s, but the little $1 \times 1$ can't do so, as it has no more columns to give. I.e., we gain a new column of 0s for each block of size at least 2.

Let's try the cube:

$$(J - \lambda I)^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & & \\ & & & & 0 & 0 & 0 & & & & & & \\ & & & & 0 & 0 & 0 & & & & & & \\ & & & & 0 & 0 & 0 & & & & & & \\ & & & & & & & 0 & 0 & & & & \\ & & & & & & & 0 & 0 & & & & \\ & & & & & & & & & 0 & 0 & & \\ & & & & & & & & & 0 & 0 & & \\ & & & & & & & & & & & 0 \end{pmatrix}$$

This now has eleven columns of 0s. The two new columns appeared in the two biggest blocks, because this time the $2 \times 2$ blocks have no more columns to give. So this time, we gain a new column of 0s for each block of size at least 3.

And the fourth power:

$$(J - \lambda I)^4 = \begin{pmatrix} 0 & 0 & 0 & 0 & & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & & \\ & & & & 0 & 0 & 0 & & & & & & \\ & & & & 0 & 0 & 0 & & & & & & \\ & & & & 0 & 0 & 0 & & & & & & \\ & & & & & & & 0 & 0 & & & & \\ & & & & & & & 0 & 0 & & & & \\ & & & & & & & & & 0 & 0 & & \\ & & & & & & & & & 0 & 0 & & \\ & & & & & & & & & & & 0 \end{pmatrix}$$

This time, only the biggest block can give us a new column of 0s, because the $3 \times 3$ block ran out.

We see that as we go from $(J - \lambda I)^{k-1}$ to $(J - \lambda I)^k$ increases by 1 for each block still able to provide us with a column of 0s. And since in $(J - \lambda I)^{k-1}$, any block of size less than $k$ is already entirely 0, we see that this increase is precisely the number of blocks of size at least $k$.

In notation, we have $n((J - \lambda I)^k) - n((J - \lambda I)^{k-1}) =$ number of $\lambda$-blocks of size at least $k$.

This formula holds for $A$ as well as $J$. Since $J$ is the JNF of $A$, there is some invertible matrix $P$ with $J = P^{-1}AP$. Then $J - \lambda I = P^{-1}AP - \lambda I = P^{-1}(A - \lambda I)P$, and more generally $(J - \lambda I)^k = P^{-1}(A - \lambda I)^k P$, and since $P$ is invertible we have $n((J - \lambda I)^k) = n((A - \lambda I)^k)$.

So the number of blocks of each size in the JNF is determined by $A$, and hence the JNF is unique.

In this example, whatever the original $A$ was, we'd find:

$n(A - \lambda I) = 5$, so there are 5 $\lambda$-blocks in total

$n((A - \lambda I)^2) - n(A - \lambda I) = 4$, so there are 4 blocks of size $\geqslant 2$, and hence $5 - 4 = 1$ block of size 1.

$n((A - \lambda I)^3) - n((A - \lambda I)^2) = 2$, so there are 2 blocks of size $\geqslant 3$, and hence $4 - 2 = 2$ blocks of size 2

$n((A - \lambda I)^4) - n((A - \lambda I)^3) = 1$, so there is 1 block of size $\geqslant 4$, and hence $2 - 1 = 1$ block of size 3.

$n((A - \lambda I)^5) - n((A - \lambda I)^4) = 0$, so there are no blocks of size $\geqslant 5$, and hence $1 - 0 = 1$ block of size 4.

So the number of blocks of each size is determined, and hence so is the JNF. (This method works for matrices with more than one eigenvalue, of course.)