

Minimal Polynomials and Jordan Normal Forms

1. These polynomials are small; those polynomials are far away

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 $I, A, A^2, \dots, A^{n^2}$. There are $n^2 + 1$ elements here, so they are linearly dependent: $a_0 I + a_1 A + \dots + a_{n^2} A^{n^2} = 0$ for some a_i , not all 0. Let $p(t) = \sum a_i t^i$, 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 $\deg(p) \geq \deg(m)$. 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 λ 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 $k \geq 0$. 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 $m(\lambda) = 0$.

M5. $\chi(t)$ and $m(t)$ are invariant under change of basis.

Proof. Let $B = P^{-1}AP$. Since $B^n = (P^{-1}AP)^n = P^{-1}A^n P$, 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 $\chi_B(t) = \chi_A(t)$.

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

Proof. See the notes.

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

M6. If $\chi(t) = \prod_{i=1}^n (t - \lambda_i)^{a_i}$, what are the possibilities for $m(t)$?

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

2. Prisoner: Cell Block J

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 λ , 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) λ s. The actual details of what sizes and λ 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.)

J1. Let J be a Jordan block with diagonal entries λ . Then λ 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 λ . 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 $x_2 = \dots = x_n = 0$, giving the eigenvector $(1, 0, \dots, 0)$.

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

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

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.

Example. Knowing χ and m isn't enough if A is a 4×4 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 λ . 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 4×4 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 6×6 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 7×7 matrices which have the same characteristic and minimal polynomials, and the same eigenspace dimensions, but which have distinct JNFs.

4. "These look great! Where can I find one?"

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 λ , then we could pick any non-zero e_n that isn't in $\ker(A - \lambda I)^{n-1}$. Then we can let $e_{n-1} = (A - \lambda I)e_n$, $e_{n-2} = (A - \lambda I)e_{n-1}, \dots, e_1 = (A - \lambda I)e_2$. 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.