Minimal Polynomial and Jordan Form

Tom Leinster

The idea of these notes is to provide a summary of some of the results you need for this course, as well as a different perspective from the lectures.

Minimal Polynomial

Let V be a vector space over some field k, and let $\alpha : V \longrightarrow V$ be a linear map (an 'endomorphism of V'). Given any polynomial p with coefficients in k, there is an endomorphism $p(\alpha)$ of V, and we say that p is an *annihilating polynomial* for α if $p(\alpha) = 0$.

Our first major goal is to see that for any α, the annihilating polynomials can easily be classified: they're precisely the multiples of a certain polynomial m_α. We reach this goal on the next page.

Let us say that a polynomial m is *a minimal polynomial for* α (note: 'a', not 'the') if:

- $m(\alpha) = 0$
- if p is any non-zero polynomial with $p(\alpha) = 0$, then $\deg(m) \leq \deg(p)$
- the leading coefficient of m is 1 (i.e. m is monic).

(We don't assign a degree to the zero polynomial. Note that the last condition implies $m \neq 0$.)

We'd like to be able to talk about *the* minimal polynomial of α, and the following result legitimizes this:

Proposition 1 *There is precisely one minimal polynomial for α.*

Sketch Proof *Existence:* First prove that there exists a non-zero annihilating polynomial, as in your notes (using the fact that the space of endomorphisms on V is finite-dimensional). There is then a non-zero annihilating polynomial M of least degree, and if c is the leading coefficient of M then $\frac{1}{c}M$ is a minimal polynomial.

Uniqueness: Suppose that m and m' are different minimial polynomials. Then $m - m'$ is non-zero by assumption, and is an annihilating polynomial. But m and m' have the same degree, and each has leading coefficient 1, so $m - m'$ has degree less than that of m. This contradicts minimality of the degree of m. $\square$

I will write m_α for the minimal polynomial of α. More important than the fact that it has minimal degree is this result (our 'first major goal'):

Proposition 2 *For any polynomial p, $p(\alpha) = 0 \Leftrightarrow m_\alpha | p$.*

Proof $\Leftarrow$ is easy. For $\Rightarrow$, we use the result you might know as 'Euclid's algorithm' or 'the division algorithm': there are polynomials q and r such that $p = q \cdot m_\alpha + r$ and r is either 0 or has degree less than that of m_α. Now,

$$r(\alpha) = p(\alpha) - q(\alpha) \cdot m_\alpha(\alpha) = 0 - 0 = 0,$$

so by definition of minimal polynomial, $r = 0$. Hence $m_\alpha | p$. $\square$

You might compare this result to a similar one about least common multiples: if a and b are two natural numbers and l their least common multiple, then what matters most about l is not that it's 'least' in the literal sense of the word, but that a number is a common multiple of a and b if and only if it is a multiple of l. In other words, the common multiples are precisely the multiples of l.

An example of Proposition 2 in action: the Cayley-Hamilton Theorem says that $\chi_\alpha(\alpha) = 0$, where χ_α is the characteristic polynomial of α, and so we conclude that $m_\alpha | \chi_\alpha$.

From now on, let's assume that the field k we are working over is the field of complex numbers, $\mathbb{C}$. This makes life much easier, because all polynomials split into linear factors. In particular, we can write

$$\chi_\alpha(t) = \pm(t - \lambda_1)^{r_1} \cdots (t - \lambda_k)^{r_k}$$

where $\lambda_1, \ldots, \lambda_k$ are distinct scalars and $r_i \geq 1$ for each i. Of course, $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of α.

The big theorem concerning minimal polynomials, which tells you pretty much everything you need to know about them, is as follows:

Theorem 3 *The minimal polynomial has the form*

$$m_\alpha(t) = (t - \lambda_1)^{s_1} \cdots (t - \lambda_k)^{s_k} \qquad (*)$$

for some numbers s_i with $1 \leq s_i \leq r_i$. Moreover, α is diagonalizable if and only if each $s_i = 1$.

Lots of things go into the proof. Let's take it step by step.

First of all, we know the Cayley-Hamilton Theorem (itself quite a large result); and as observed above, this tells us that m_α has the form $(*)$ with $s_i \leq r_i$.

Secondly, we need to see that each $s_i \geq 1$, i.e. that every eigenvalue is a root of the minimal polynomial. So, let $i \in \{1, \ldots, k\}$ and let v be a λ_i-eigenvector. For any polynomial p, we have $p(\alpha)v = p(\lambda_i)v$, and in particular this holds for $p = m_\alpha$: hence $m_\alpha(\lambda_i)v = 0$. But v is nonzero (being an eigenvector), so $m_\alpha(\lambda_i) = 0$, as required.

Thirdly, suppose we know that α is diagonalizable. This means it has a basis of eigenvectors, whose eigenvalues are $\lambda_1, \dots, \lambda_k$, and it's easy to calculate that

$$(t - \lambda_1) \cdots (t - \lambda_k)$$

is an annihilating polynomial for α. So this is the minimal polynomial.

Finally, then, we have to show that if all the s_i's are 1 (i.e. the minimal polynomial splits into distinct linear factors) then α is diagonalizable. There's a proof of this which I like and I believe is different to the one in your notes, so I'll write it out in full. It goes via a lemma:

Lemma 4 *If* $U \xrightarrow{\beta} V \xrightarrow{\gamma} W$ *are finite-dimensional vector spaces and linear maps, then*

$$\dim \mathrm{Ker}(\gamma \circ \beta) \leq \dim \mathrm{Ker}(\gamma) + \dim \mathrm{Ker}(\beta).$$

Proof Observe that $\mathrm{Ker}(\gamma \circ \beta) = \beta^{-1}(\mathrm{Ker}(\gamma))$. (By using '$\beta^{-1}$' I am *not* suggesting that β is invertible; this is the notation for the pre-image or inverse image of a subset under a function, as explained in Numbers and Sets.)

Now, consider the function

$$\begin{array}{rccc} \beta' : & \beta^{-1}(\mathrm{Ker}(\gamma)) & \longrightarrow & \mathrm{Ker}(\gamma), \\ & u & \longmapsto & \beta(u). \end{array}$$

Applying the rank-nullity formula we get

$$\dim \beta^{-1}(\mathrm{Ker}(\gamma)) = \dim \mathrm{Im}(\beta') + \dim \mathrm{Ker}(\beta'),$$

and adding to this our initial observation and the facts that $\mathrm{Im}(\beta') \leq \mathrm{Ker}(\gamma)$ and $\mathrm{Ker}(\beta') \leq \mathrm{Ker}(\beta)$, the lemma is proved. $\square$

The hard work is now done. Supposing that all the s_i are 1, the composite of the maps

$$V \xrightarrow{\alpha - \lambda_k I} V \xrightarrow{\alpha - \lambda_{k-1} I} \cdots \xrightarrow{\alpha - \lambda_1 I} V$$

is 0. So

$$\begin{aligned} \dim(V) &= \dim \mathrm{Ker}((\alpha - \lambda_1 I) \circ \cdots \circ (\alpha - \lambda_k I)) \\ &\leq \dim \mathrm{Ker}(\alpha - \lambda_1 I) + \cdots + \dim \mathrm{Ker}(\alpha - \lambda_k I) \\ &= \dim(\mathrm{Ker}(\alpha - \lambda_1 I) \oplus \cdots \oplus \mathrm{Ker}(\alpha - \lambda_k I)), \end{aligned}$$

where the inequality comes from the Lemma (and an easy induction), and the second equality is justified by the fact that the sum of the eigenspaces is a *direct* sum. Hence the sum of the eigenspaces has the same dimension as V, i.e. this sum *is* V, and α is diagonalizable.

Invariants

When two square matrices are conjugate, you ought to think of them as essentially the same. It's just a matter of change of basis: if $B = P^{-1}AP$ then A and B are 'the same but viewed from a different angle'. (The change of basis, or change of perspective, is provided by P.)

A similar point is that the matrix of a linear map depends on a choice of basis; a different choice of basis gives a different but conjugate matrix. So, for instance, if we want to define the trace of a linear endomorphism as the trace of the matrix representing it, then in order for this to make sense we have to check that

$$\mathrm{trace}(P^{-1}AP) = \mathrm{trace}(A)$$

for any $n \times n$ matrices A, P with P invertible. A fancy way of putting this is 'trace is invariant under conjugation'. The things which are invariant under conjugation tend to be the things which are important.

So, what are the interesting 'invariants' of linear endomorphisms (or square matrices, if you prefer)? Certainly:

- the eigenvalues
- their algebraic multiplicities
- their geometric multiplicities (i.e. the dimensions of the eigenspaces)
- the characteristic polynomial
- the minimal polynomial.

(In fact, the characteristic polynomial tells you exactly what the eigenvalues and algebraic multiplicities are, so it wasn't really necessary to mention them separately. Recall that the algebraic multiplicity of an eigenvalue λ is the power of $(t - \lambda)$ occurring in $\chi_\alpha(t)$.) To this list we might also add:

- the trace
- the determinant
- the rank
- the nullity.

However, these four are already covered by the first list. The trace of an endomorphism α of an n-dimensional space is

$$(-1)^{n-1}.(\text{coefficient of } t^{n-1} \text{ in } \chi_\alpha(t)).$$

The determinant of α is $\det(\alpha - 0.I) = \chi_\alpha(0)$. The rank is n minus the nullity; and the nullity is $\dim \mathrm{Ker}(\alpha - 0.I)$, which is 0 if 0 is not an eigenvalue, and is the geometric multiplicity of 0 if it is an eigenvalue.

Moreover, a look at the minimal polynomial tells you at a glance whether the matrix (or map) is diagonalizable—another important property, again invariant under conjugation.

So, the conclusion is that the characteristic polynomial, minimal polynomial and geometric multiplicities tell you a great deal of interesting information about a matrix or map, including probably all the invariants you can think of. Usually it takes an appreciable amount of work to calculate these invariants for a given matrix. In the next section, we'll see that for a matrix in Jordan canonical form they can be read off instantly.

Be warned that the invariants I've mentioned don't tell you everything: there exist pairs of matrices for which all these invariants are the same, and yet the matrices are not conjugate. An example is given in the next section.

Jordan Canonical Form

In this section I'll mostly work with matrices rather than linear maps, just for a change. You should be fairly happy about switching between the two modes of thought.

I'll state the main theorem (the proof of which is off the syllabus), then I'll try to explain what the point is and how the finer details work.

First we need some notation. Let $A_1, \ldots, A_m$ be square matrices, and let

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_m \end{pmatrix}. \tag{$\dagger$}$$

The square matrices A_i can be of different sizes; thus if A_i is a $n_i \times n_i$ matrix then A is a $n \times n$ matrix, where $n = n_1 + \cdots + n_m$. The entries '0' denote zero matrices of the appropriate sizes. For convenience, I will write the matrix on the right-hand side of (†) as

$$A_1 \oplus \cdots \oplus A_m.$$

For any $d \geq 1$ and complex number λ, let $J_\lambda^{(d)}$ be the $d \times d$ matrix

$$J_\lambda^{(d)} = \begin{pmatrix} \lambda & 1 & & & & & \\ & \lambda & 1 & & & & \\ & & \lambda & 1 & & & \\ & & & \ddots & \ddots & & \\ & & & & \lambda & 1 & \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{pmatrix}$$

where all the unmarked entries are 0. This is a so-called *Jordan block*. (Note that when $d = 1$, it is just the 1×1 matrix (λ).)

The big theorem is:

Theorem 5 *Let A be a square matrix of complex numbers. Then there are natural numbers $m, n_1, \ldots, n_m \geq 1$ and complex numbers $\mu_1, \ldots, \mu_m$ such that A is conjugate to*

$$J_{\mu_1}^{(n_1)} \oplus \cdots \oplus J_{\mu_m}^{(n_m)}. \qquad (\ddagger)$$

Moreover, if A is also conjugate to

$$J_{\mu'_1}^{(n'_1)} \oplus \cdots \oplus J_{\mu'_{m'}}^{(n'_{m'})}$$

then $m' = m$ and there is a permutation $\sigma \in S_m$ such that for all i, we have $n'_i = n_{\sigma(i)}$ and $\mu'_i = \mu_{\sigma(i)}$.

You are not required to know the proof. A matrix of the form ($\ddagger$) is said to be in *Jordan canonical form*, or *Jordan normal form*. The 'moreover' part says that the Jordan canonical form of a matrix is as unique as it possibly could be: that is, unique up to permutation of the blocks. There's no way it could be *genuinely* unique, since for any square matrices C and D (perhaps of different sizes), the two matrices $C \oplus D$ and $D \oplus C$ are conjugate.

So, what's the point of the Jordan canonical form? Here are three answers:

An approximation to diagonalizability If a square matrix or linear endomorphism can be put into diagonal form then it is very easy to work with, and one might at first hope that *every* matrix is diagonalizable. But this isn't true. An easy example is

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

its characteristic polynomial is t^2, so its only eigenvalue is 0; but the eigenvalues of a diagonal matrix are the entries on the diagonal, so the only diagonal matrix it might be conjugate to is 0. However, the only matrix conjugate to 0 is 0 itself.

So not every matrix is diagonalizable. But we do know that every complex matrix (square, as usual) is triangularizable, which is some sort of approximation. The Jordan theorem is a much, much more incisive statement than this. Of course, every matrix in Jordan canonical form is upper triangular; in fact, it looks like

$$\begin{pmatrix} * & \# & & & \\ & * & \# & & \\ & & \ddots & \ddots & \\ & & & * & \# \\ & & & & * \end{pmatrix}$$

where the entries labelled $*$ are any complex numbers, the entries labelled $\#$ are either 0 or 1, and all the unlabelled entries are 0. (But not every matrix of this form is in Jordan canonical form: exercise.) In summary, we can view the Jordan theorem as being the next-best thing to the statement 'every matrix is diagonalizable', and it has the advantage of being true.

A convenient way of displaying invariants Earlier we discussed the important 'invariants' of square matrices: those quantities which don't change under conjugation (and can therefore be assigned to linear endomorphisms). All the ones we (probably) know about can, as we saw, be derived from the characteristic polynomial, the minimal polynomial and the geometric multiplicities of the eigenvalues. Given a matrix in Jordan canonical form, these can be read off immediately, without any need for calculation. I'll show you how to do this in a moment.

Like prime factorization Any natural number $n \geq 1$ is equal to $p_1 p_2 \cdots p_m$ for some sequence $p_1, \ldots, p_m$ of prime numbers; this sequence is unique, but for the obvious fact that the prime factors can be permuted. This statement has a clear resemblance to the Jordan theorem, where the Jordan blocks $J_\lambda^{(d)}$ play the role of the prime numbers. They are the 'building blocks' of square matrices, just as the primes are the building blocks of natural numbers.

This provides a neat way to think about the theorem. In fact, there is a very general theorem which has as special cases both the Jordan theorem and the unique prime factorization of natural numbers, but that's way beyond this course. (It's a result from the theory of rings and modules: namely, the classification of finitely generated modules over a principal ideal domain.)

I promised to show you how lots of interesting invariants can be read off immediately from a matrix in Jordan canonical form. I'll lead you up to this via a sequence of exercises:

Exercise 1 Show that for any $d \geq 1$ and $\lambda \in \mathbb{C}$, the matrix $J_\lambda^{(d)}$ has:

a. λ as its only eigenvalue

b. minimal polynomial $(t - \lambda)^d$

c. characteristic polynomial $(t - \lambda)^d$, and that

d. the geometric multiplicity of λ is 1.

Exercise 2 (Shorter than it looks.) Let C and D be any square matrices.

a. Show that

$$\{\text{eigenvalues of } C \oplus D\} = \{\text{eigenvalues of } C\} \cup \{\text{eigenvalues of } D\}.$$

b. Show that for any polynomial p, $p(C \oplus D) = 0$ if and only if $p(C) = 0$ and $p(D) = 0$, and deduce that $m_{C\oplus D}|p$ if and only if $m_C|p$ and $m_D|p$. (In an appropriate sense, $m_{C\oplus D}$ is the least common multiple of m_C and m_D.) Then deduce that for any complex number λ,

$$s'' = \max\{s, s'\}$$

where s'' is the power of $(t - \lambda)$ occurring in $m_{C\oplus D}(t)$ (when it's been written as a product of linear factors), and similarly s for m_C and s' for m_D.

c. Show that $\chi_{C\oplus D}(t) = \chi_C(t).\chi_D(t)$. Deduce that for any complex number λ,

$$r'' = r + r'$$

where r, r' and r'' are defined as in the previous part of the question, but with the characteristic polynomial instead of the minimal polynomial.

d. Show that for any complex number λ,

$$g'' = g + g'$$

where g'' is the geometric multiplicity of λ in $C \oplus D$, and similarly g for C and g' for D. (If λ is not an eigenvalue of C then g is by definition 0: so in any case, $g = \dim \operatorname{Ker}(C - \lambda I)$. The same convention applies to g' and g''.)

Exercise 3 Let λ be a complex number, let $g, d_1, \ldots d_g$ be natural numbers, and let $A = J_\lambda^{(d_1)} \oplus \cdots \oplus J_\lambda^{(d_g)}$. Show that for the matrix A,

a. the only eigenvalue is λ

b. the power of $(t - \lambda)$ in m_A is $\max\{d_1, \ldots, d_g\}$

c. the power of $(t - \lambda)$ in χ_A is $d_1 + \cdots + d_g$

d. the geometric multiplicity of λ is g.

Exercise 4 Let $\lambda_1, \ldots, \lambda_k$ be complex numbers, let $d_1^1, \ldots, d_1^{g_1}, \ldots, d_k^1, \ldots d_k^{g_k}$ be natural numbers (superscripts don't indicate powers), and let

$$A = (J_{\lambda_1}^{(d_1^1)} \oplus \cdots \oplus J_{\lambda_1}^{(d_1^{g_1})}) \oplus \cdots \oplus (J_{\lambda_k}^{(d_k^1)} \oplus \cdots \oplus J_{\lambda_k}^{(d_k^{g_k})}).$$

Show that for the matrix A,

a. the eigenvalues are $\lambda_1, \ldots, \lambda_k$

b. the power of $(t - \lambda_i)$ in m_A is $\max\{d_i^1, \ldots, d_i^{g_i}\}$

c. the power of $(t - \lambda_i)$ in χ_A is $d_i^1 + \cdots + d_i^{g_i}$

d. the geometric multiplicity of λ_i is g_i.

Let's re-express the results of this final exercise. It says that given a matrix in Jordan canonical form,

- the eigenvalues are the entries down the diagonal
- $m_A(t) = (t-\lambda_1)^{s_1} \cdots (t-\lambda_k)^{s_k}$ where s_i is the size of the largest λ_i-block in A (and the λ_i's are the distinct eigenvalues)
- $\chi_A(t) = (t-\lambda_1)^{r_1} \cdots (t-\lambda_k)^{r_k}$ where r_i is the number of occurrences of λ_i on the diagonal
- the geometric multiplicity of λ_i is the number of λ_i-blocks in A.

This covers all the major theory, and if you've got this far then you can be pleased with yourself. Here are are a couple of further points.

Some of you asked me for an efficient method of calculating the Jordan form of a given matrix. The bad news is that there isn't one that I know of. The good news, however, is that for square matrices of size 6×6 or less, you can work out the Jordan form by calculating the characteristic and minimal polynomials and the dimensions of the eigenspaces. For example, suppose you're given a 6×6 matrix and you calculate that its characteristic polynomial is $(t-3)^4(t-i)^2$, that its minimal polynomial is $(t-3)^2(t-i)^2$, that the 3-eigenspace is 3-dimensional, and that the i-eigenspace is 1-dimensional. Then the diagonal is made up of 4 copies of 3 and 2 of i. There are 3 Jordan 3-blocks, the largest of which is 2×2, which means that their sizes are 2, 1 and 1 (since they add up to 4). Similarly, there is just one Jordan i-block, which is 2×2. So the Jordan canonical form of the matrix is

$$\begin{pmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & i & 1 \\ 0 & 0 & 0 & 0 & 0 & i \end{pmatrix}.$$

As a rather tedious exercise, you can show that this line of reasoning will always tell you exactly what the Jordan canonical form of a square matrix is when it is at most 6×6. However, it doesn't always work for larger matrices. For example, suppose you're told that a 7×7 matrix has characteristic polynomial t^7, minimal polynomial t^3, and that the dimension of the 0-eigenspace is 3. Then the matrix could be conjugate to either one of the Jordan canonical forms

$$J_0^{(3)} \oplus J_0^{(3)} \oplus J_0^{(1)}, \qquad J_0^{(3)} \oplus J_0^{(2)} \oplus J_0^{(2)},$$

which are not conjugate to each other (by the uniqueness part of the Jordan theorem).

The moral of this last example is that if you have two $n \times n$ matrices in front of you, then even if all the invariants you can think of are the same for each (e.g. eigenvalues, rank, nullity, characteristic polynomial, trace, determinant, minimal polynomial, geometric multiplicities), it does *not* follow that the matrices are conjugate to each other.