

Phil Lucht's

JORDAN FORM BINDER

Tabs:

1

2

3

4

5

6

Front: Overview

1. The N-space drawing and notes on same, relates to page 65 of Matthews notes.

2. "PhL on Jordan form" -- some summary notes on J and C matrix, various details.

3. "EE263 Notes" (and download pages) on Jordan Form (Boyd, Stanford).

4. Notes on how to compute a Jordan form, + download (in Mathematica).

5. "Tom Leinster Notes" (and download pages) on Jordan Form.

6. "PlanetMath proof of Jordan form" notes and download pages.

~~7. Matthews Lectures, PDF list, index and then my NOTES on all these lectures.~~

~~8. Matthews Chapter 1: Linear Transformations~~

~~9. Matthews Chapter 2: Polynomials over a Field, min polys, Companion, direct sums, Galois~~

A ~~10. Matthews Chapter 3: Invariant Subspaces (T-invariant, T-cyclic, finding the min poly, PDT)~~

B ~~11. Matthews Chapter 4: The Jordan Form, dot diagram, examples, Markov, real Jordan form.~~

How I got to this subject. I was reading Jim's papers and saw mention of Fortran routines to get eigenvalues of a tridiagonal matrix by some strange QL method. This led me to do a broad review of "matrices", a piece of equipment long idle on my shop floor. When reading M&M, I saw their discussion of getting an arbitrary matrix to "triangular block diagonal form" which I called Pivot and Twist, and it was a torturous discussion. I later learned that something called the Schur Decomposition gets you by similarity to a completely triangular form, so at least this result was consistent with the M&M discussion. But things were still very hazy. I read the Fortran book about how you actually compute eigenvalues and eigenvectors numerically using the QR iteration method, a wonder of the world. But in the theory world, there is more going on here. Jordan did Jordan Form around 1870, see his bio.

Camille Jordan, by the way, was French 1838-1922, an engineer and mathematician.

(See <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Jordan.html>)

An arbitrary square matrix can be brought by similarity to Jordan Canonical Form. This form is triangular block diagonal as M&M said, but a lot more is known. It is really a double direct sum. The outer direct sum puts the eigenmanifolds into direct sum (block diagonal), and this is now proved from the Primary Decomposition Theorem which KR Matthews derives in his notes after LOTS of preparation.

Then within each eigenmanifold, there is further direct sum action. You end up with a set of small blocks with λ on diagonal, and only 1's on first off diagonal. Within each eigenmanifold, we have these facts:

- (1) the total number of diagonal elements is of course the algebraic multiplicity (sum of block dims)
- (2) the number of Jordan blocks is the geometric multiplicity
- (3) the size of the largest block is the minimal polynomial exponent b for this eigenmanifold.

In my basic study of matrices, the notion of a "minimal polynomial" never came up. I now know what it is. It has exactly the same form as the characteristic polynomial, but the exponents can all be lower. I know exactly how to compute this polynomial (in principle) for any matrix A .

In order to "fill out" the Jordan structure of a particular manifold, you need three numbers: algmult, geomult and the min poly exponent. In some cases, you don't need all three, but in general you do need all three. You can see this by thinking about the case where $\text{algmult} = 4$. I think beyond 6×6 you need even more information to get the form figured out.

When the geomult is full, you get all 1×1 blocks and you are fully diagonal, no surprise. When geomult is only 1, then you get a single full square block. In general, each square Jordan block does have a set of "generalized eigenvectors" that I know how to calculate from my EE263 notes. For each block, only one of these (the "first column") is a true eigenvector, the others are sort of fillers.

Just as with a fully diagonalizable matrix, the similarity matrix which takes you to Jordan form consists of the eigenvectors along with these filler generalized eigenvectors. Thus, you can in fact build the similarity

you need to get any matrix A into Jordan Canonical Form. I had wondered earlier how this worked when you did not have enough eigenvectors, I thought maybe some repeated. Now I know the answer.

The Jordan form is "as far as you can go". Those blocks are the primitives. Your distance from non-diagonalness is reflected by how many ones you have on that first off diagonal.

Just as a reminder, if the matrix A has no "defects", meaning $\text{geo} = \text{alg}$ in each manifold, then the Jordan form will be fully diagonal. We see now why this has to be true if all eigenvalues are different!

The authors like to make a parallel between these Jordan blocks and prime numbers:

A = sum of irreducible Jordan blocks

N = product of powers of prime numbers

In each case, you have reduced something to the most basic primitive elements.

Conclusion: I think I know how to take any matrix to Jordan form now, I could do it manually. But Maple has a simple call `jordan(A,'P')` which computes the form *and* the similarity which gets you there. I suspect this program computes the basic eigenvectors, then computes the generalized eigenvectors, then has the similarity, and then just uses it to get J .

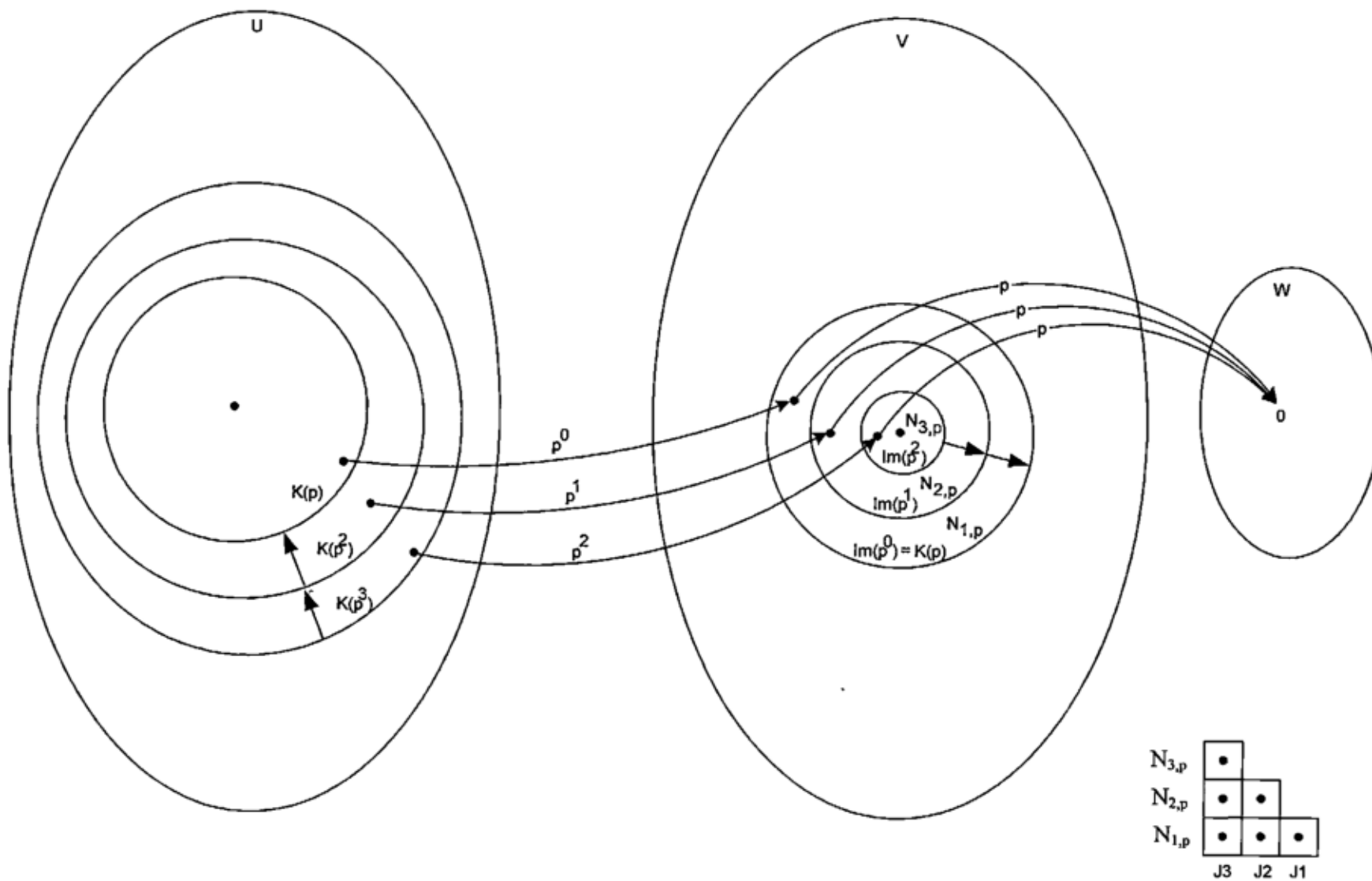


Figure 1

See Figure 1 below. Just explaining what this is a picture of is going to take a lot of words! We are talking about composite mappings of the form $p^{h-1}(T): U \rightarrow V$ then $p(T): V \rightarrow W$, for a total mapping that has the form $p^h(T): U \rightarrow W$. The left egg is U , the middle egg is V , and we only show the zero point of the right egg W . We are interested in $h = 1, 2, 3, \dots, b$ ($b = \text{min poly factor exponent}$). Our picture is drawn for $b = 3$, and we show the three h values $h = 1, 2$ and 3 . We are always interested on the left in $\text{Ker}(p^h)$. We know that $\dim \text{Ker}(p^h) = a$, the algmult. And we know that $\dim \text{Ker}(p) = \gamma$, the geomult. We know from Matthews page 58 that the nullspaces in the left egg are nested as shown, without equality. In other words, we know that $\text{Ker}(p^3) \supset \text{Ker}(p^2) \supset \text{Ker}(p)$. So this sequence of kernel spaces is getting smaller, while the corresponding sequence of image spaces in the V egg is getting larger. Think of nullity + rank = constant as the explanation for this fact. As nullity goes down (smaller spaces on the left), the rank goes up (size of domain on the right).

As an example of what this picture shows, consider a point in $\text{Ker}(p^3)$ like the one shown as $\bullet$. We operate on this point with $p^2(T)$ and we end up with a point in the space we have labeled $\text{Im}(p^2)$. All of this region is then going to map under the final p operator into 0 in the W egg. Now it may be that some points in $\text{K}(p^3)$ map under p^2 directly into the 0 in the V egg, but OK, the final p mapping will still take us to 0 in the W egg. All points in $\text{K}(p^3)$ under the combined mapping p^2 then p must end up at 0 in W . But they end up in an intermediate zone in V which we have called $\text{Im}(p^2)$.

Now we have a bit of confusion. When we say $\text{Im}(p^2)$ here and in the picture, we really mean the image or range of the operator Matthews calls Q_0 , not the operator Q . Q_0 is domain-restricted to one of the kernels on the left. For example, we might have $Q = p^2$ with domain being all of U , but $Q_0 = p^2$ with domain being only $\text{Ker}(p^3)$. If we did not do this restriction, the image areas in the V egg would be larger, and in fact they would probably have regions which lie outside the largest circle in the V egg which is in fact $\text{Ker}(p)$. But I only care about these "restricted" image regions, because I only care about starting off in one of the kernel regions on the far left.

Because my images in the V egg are for the "restricted" operators, I think we can identify each of these regions with one of Matthew's $N_{h,p}$ spaces. Normally we are supposed to think of an N -space as the intersection (what I called a cross-hatched region elsewhere) of an Image area with $\text{Ker}(p)$ in the V egg, but with this restricted meaning of Image, the nested set of images all lie inside $\text{Ker}(p)$ as shown, so we don't have to think about intersection. The outermost region in the V egg is $N_{1,p} = \text{Ker}(p)$. [At least I think this is all correct, I could be wrong.]

So to summarize: on the left we have a set of nullspaces which are *decreasing* in size, and on the right we have a set of N -spaces that are *increasing* in size and which are maxing out at $N_{1,p} = \text{ker}(p)$. When we say "increasing in size", we mean that we are adding one or more new basis vectors. For example, suppose $N_{3,p}$ has 1 basis vector and $N_{2,p}$ has 2 and $N_{1,p}$ has 3. In Matthews notation, this means $v_{3,p} = 1$ and $v_{2,p} = 2$ and $v_{1,p} = 3$. This is in fact how we get the Matthew's dot diagram shown in our figure. As you move outward to each larger N space, Matthews suggests that you "re use" the basis vectors you already have and add new ones as needed to this set. This is his notion of "extending" the basis.

Now let's re-examine Matthew's example on page 73 which in fact has the dot diagram shown, and we will use his other numbers in this discussion as well. We are suppose to "start off" at the top of the dot diagram, which is to say, we start thinking about the smallest nested N space which is $N_{3,p}$. We want to find basis vectors for this space. His suggestion for doing this is as follows:

(1) *first*, find basis vectors over on the left for space $\text{Ker}(p^3)$. In his example, $\dim \text{Ker } p^3 = 6 = a$, so there are in fact 6 basis vectors. To mechanically do this of course requires significant work, but we

would have Maple do that work for us. We would compute that matrix $p^3(A) = (A - \lambda I)^3$ [it happens that $\lambda = 0$ in his example], and we would then ask Maple to find the eigenvectors of this matrix. In his example, this is a single eigenmanifold with $\lambda = 0$, and there really will be 6 distinct eigenvectors since we know that $\dim \text{Ker } p^3 = 6$. Matthews calls these eigenvectors X_1 through X_6 .

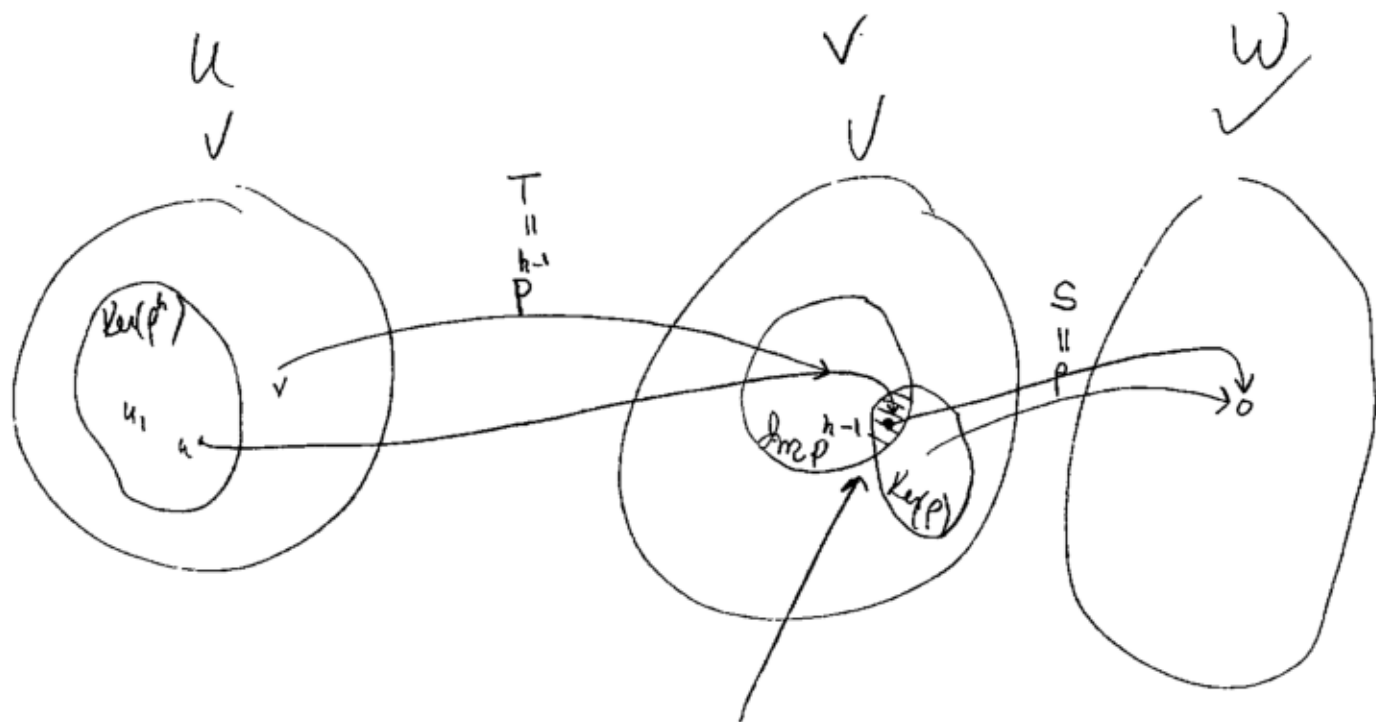
(2) *second*, Matthews suggests that you apply p^2 to all the basis vectors that you found for $\text{Ker } p^3$ and in this way you will generate a set of vectors which "span" $N_{3,p}$. He uses a certain bracket $\langle \dots \rangle$ notation to indicate a set of spanning vectors, so here he would say $N_{3,p} = \langle p^2 X_1, p^2 X_2, p^2 X_3, \dots, p^2 X_6 \rangle$ and this notation means "the space $N_{3,p}$ is spanned by the vectors $\dots$ ". Of course we know that $v_{3,p} = 1$, so that tells us that this set of spanning vectors contains only one LI vector! Perhaps several of them are 0 (under the action of p^2), and then any remaining ones must be "multiples" of each other. Could we get $p^2 X_2 = 0$, for example? Sure! This would just mean that the basis vector X_2 of $\text{Ker } p^3$ is also a basis vector of $\text{Ker } p^2$. You could imagine creating a whole world of basis vectors over in the left egg by a similar extension idea. In any event, we know that the set shown above will have a viable basis vector for $N_{3,p}$. Let's assume it is $p^2 u_1$. Then we can simplify to say that $N_{3,p} = \langle p^2 u_1 \rangle$.

Matthews then moves on to think about basis vectors for the next larger space $N_{2,p}$. He re-uses the one we just found, which we have called $p^2 u_1$. He then computes a set of basis vectors (BV's) for $\text{Ker } p^2$ (and we know there are 5 of these), and then says $N_{2,p} = \langle p^2 u_1, pY_1, pY_2, pY_3, \dots, pY_5 \rangle$ where he prepends our already-found BV. We know that $v_{2,p} = 2$, so one of the pY_i guys will be our pu_2 second BV, and we will then say $N_{2,p} = \langle p^2 u_1, pu_2 \rangle$. Notice that when we worked with $\text{Ker } p^3$, the mapping operators were p^2 . And now when we are dealing with $\text{Ker } p^2$, the mapping operators are p .

We then go the last step and we end up with $N_{1,p} = \langle p^2 u_1, p^1 u_2, p^0 u_3 \rangle$, and this set of BV's works for our $\text{Ker}(p)$ space which has $v_{1,p} = 3 = \text{geomult}$.

Now we ask: WHY does Matthews suggest we keep re-using previously computed BV's and adding new ones to extend as needed? After all, we could do it some other way. The reason is this: Look at our result $N_{1,p} = \langle p^2 u_1, p^1 u_2, p^0 u_3 \rangle$. The first BV is not only a BV for $\text{Ker}(p)$, but it is also the "maximal" BV of a cyclic subspace C that is associated with the Jordan block. The BV's for this C space are $\langle p^0 u_1, p^1 u_1, p^2 u_1 \rangle$. This C space is generated by vector u_1 and has the min poly $p^2 = p^{3-1} = p^{e-1}$. It has dimension $e = 3$, and it is associated with the first column of the dot diagram. So to restate: this BV has two roles to play. It is a BV for $\text{Ker}(p)$, and it is a BV for J_3 . Similarly, $p^1 u_2$ is a BV for $\text{Ker}(p)$, and is also a BV for J_2 .

So the plan in the above example is to find viable vectors u_1, u_2 , and u_3 in U and from these we can compute both the BV's for $\text{Ker}(p)$ as well as the matrix P which gets us to Jordan form. The "other" BV's for the C spaces that are not in our $\text{Ker}(p)$ BV list are called "generalized eigenvectors" and we can see that there is no question as to their existence. In Matthew's scheme, you can trivially compute them by applying powers of p to the starting u_i . It is all pretty automated!



= $N_{h,p}$

as h increases, $N_{h,i}$ gets smaller
 smallest is $N_{b,p}$ with $v_{b,p} \geq 1$.

$p(t)/f_i(t)/v_i$

1 1 1 1

$e_i = k_i$

$\{a_i v_i = 0$

Fig 2

Notes on Linear Operators

Theorem 1. Let $T : U \rightarrow V$ be a linear mapping. Then U is finite-dimensional iff $\text{Ker}(T)$ and $\text{Im}(T)$ are finite-dimensional in which case

$$\dim(U) = \dim(\text{Ker}(T)) + \dim(\text{Im}(T)). \quad N = \text{nullity} + \text{rank}.$$

Proof. ($\Rightarrow$) If U is finite-dimensional then so is $\text{Ker}(T)$ since a subspace of a finite-dimensional vector space is also finite-dimensional. Also, if $U = \text{Span}(e_1, \dots, e_n)$, then $\text{Im}(T) = \text{Span}(T(e_1), \dots, T(e_n))$ so that the image of T is also finite-dimensional. Now if $f_1, \dots, f_s$ is a basis for $\text{Ker}(T)$ and we complete $f_1, \dots, f_s$ to a basis $f_1, \dots, f_s, f_{s+1}, \dots, f_{s+r}$ of U then we claim that $T(f_{s+1}), \dots, T(f_{s+r})$ is a basis for $\text{Im}(T)$. Indeed, they span $\text{Im}(T)$ since $T(f_1) = \dots = T(f_s) = 0$, and if $a_1 T(f_{s+1}) + \dots + a_r T(f_{s+r}) = 0$ we have $T(a_1 f_{s+1} + \dots + a_r f_{s+r}) = 0$ which implies $a_1 f_{s+1} + \dots + a_r f_{s+r} = b_1 f_1 + \dots + b_s f_s$. Bringing all terms to the left side we get a dependence relation among $f_1, \dots, f_{s+r}$. Since f_i 's are linearly independent we get $a_1 = a_2 = \dots = a_r = 0$. This yields $\dim(U) = s + r = \dim(\text{Ker}(T)) + \dim(\text{Im}(T))$.

($\Leftarrow$) Now suppose $\text{Ker}(T)$ and $\text{Im}(T)$ are finite-dimensional. Let $f_1, \dots, f_s$ be a basis for $\text{Ker}(T)$ and let $h_1, \dots, h_r$ be a basis for $\text{Im}(T)$. We have $h_i = T(f_{s+i})$ with $f_{s+1}, \dots, f_{s+r} \in U$. We claim that $f_1, \dots, f_{s+r}$ is a basis for U . Indeed, if $u \in U$, then $T(u) = a_{s+1} T(f_{s+1}) + \dots + a_{s+r} T(f_{s+r})$ which implies that $u - a_{s+1} f_{s+1} - \dots - a_{s+r} f_{s+r} \in \text{Ker}(T)$ and hence that

$$u - a_{s+1} f_{s+1} - \dots - a_{s+r} f_{s+r} = a_1 f_1 + \dots + a_s f_s$$

which gives $u = a_1 f_1 + \dots + a_{s+r} f_{s+r}$ and hence that $f_1, \dots, f_{s+r}$ generate U . To show linear independence of these vectors suppose that $a_1 f_1 + \dots + a_{s+r} f_{s+r} = 0$. Applying T to both sides yields $a_{s+1} h_1 + \dots + a_{s+r} h_r = 0$ which gives $a_{s+1} = \dots = a_{s+r} = 0$ since $h_1, \dots, h_r$ are linearly independent. But then $a_1 f_1 + \dots + a_s f_s = 0$ which gives $a_1 = \dots = a_s = 0$ by the fact that $f_1, \dots, f_s$ are linearly independent. Thus $\dim(U) = r + s = \dim(\text{Ker}(T)) + \dim(\text{Im}(T))$. $\square$

Corollary 2. Let $T : U \rightarrow V, S : V \rightarrow W$ be linear mappings such that $\text{Ker}(S)$ and $\text{Ker}(T)$ are finite-dimensional. Then

$$\dim \text{Ker}(ST) = \dim \text{Ker}(T) + \dim(\text{Ker}(S) \cap \text{Im}(T)).$$

Proof. We first note that $\text{Ker}(T) \subseteq \text{Ker}(ST)$ and that

$$u \in \text{Ker}(ST) \iff ST(u) = 0 \iff T(u) \in \text{Ker}(S) \cap \text{Im}(T). \quad \checkmark \text{ see picture!}$$

Now let $T_0 : \text{Ker}(ST) \rightarrow W$ be the linear mapping defined by restriction of T to $\text{Ker}(ST)$. Then $\text{Ker}(T_0) = \text{Ker}(T)$ and $\text{Im}(T_0) = \text{Ker}(S) \cap \text{Im}(T)$ which yields the result. $\square$

Corollary 3. $\dim(\text{Ker}(ST)) \leq \dim(\text{Ker}(S)) + \dim(\text{Ker}(T))$ with equality if $\text{Ker}(S) \subseteq \text{Im}(T)$, in particular, if T is surjective.

Theorem 4. Let T be a linear operator on a vector space V and let $a_1, a_2, \dots, a_k$ be distinct scalars such that $\dim(\text{Ker}(T - a_i)^{n_i})$ is finite-dimensional for $1 \leq i \leq k$. Then

$$\text{Ker}((T - a_1)^{n_1} (T - a_2)^{n_2} \dots (T - a_k)^{n_k}) = \text{Ker}(T - a_1)^{n_1} + \text{Ker}(T - a_2)^{n_2} + \dots + \text{Ker}(T - a_k)^{n_k}.$$

Proof. Let $W = \text{Ker}((T - a_1)^{n_1} (T - a_2)^{n_2} \dots (T - a_k)^{n_k})$ and let $W_i = \text{Ker}(T - a_i)^{n_i}$. We first prove that $\dim(W_1 + \dots + W_k) = \dim(W_1) + \dots + \dim(W_k)$. For this it suffices to prove that if $w_i \in W_k$ with $w_1 + w_2 + \dots + w_k = 0$ then $w_1 = w_2 = \dots = w_k = 0$. Let S_i be the product of the operators

$(T - a_j)^{n_j}$ with $j \neq i$. Then, applying S_i to both sides of $w_1 + \dots + w_k = 0$, we get $S_i(w_i) = 0$ since $S_i(w_j) = 0$ for $j \neq i$. Since $S_i(W_i) \subseteq W_i$, the restriction of S_i to W_i is a linear operator on W_i and $S_i = \prod_{j \neq i} (T_j - a_j)^{n_j}$, where T_j is the restriction of T to W_j . That $w_i = 0$ follows from the following Lemma since

Lemma 5. *If S is a linear operator on a vector space W and a is a scalar such that $(S - a)^n = 0$ then $S - b$ is invertible for every $b \neq a$.*

Proof. We first prove this in the case $a = 0, b = 1$. Then $S^n = 0$ so that

$$(1 - S)(1 + S + S^2 + \dots + S^{n-1}) = 1 + S + S^2 + \dots + S^{n-1} - S - S^2 - \dots - S^{n-1} = 1.$$

Since two polynomials in S commute we get that $1 - S$ is invertible with inverse $1 + S + S^2 + \dots + S^{n-1}$. Since $S - 1 = -(1 - S)$ we see that $S - 1$ is also invertible. The general case follows from the identity $S - b = (a - b)(1 - (b - a)^{-1}(S - a))$. $\square$

Returning to the proof of Theorem 4, we have $Z = W_1 + W_2 + \dots + W_k \subseteq W$ since $W_i \subseteq W$ which implies that $\dim Z \leq \dim W$. But, by Corollary 3,

$$\dim W \leq \dim W_1 + \dim W_2 + \dots + \dim W_k = \dim Z$$

so that $\dim Z = \dim W$ and hence that $Z = W$. $\square$

Corollary 6. *If T is a linear operator on a finite-dimensional vector space, then T is diagonalizable if and only if there are distinct scalars $a_1, a_2, \dots, a_k$ such that $(T - a_1)(T - a_2) \dots (T - a_k) = 0$.*

We now apply these results to the case of the differential operator D and the left-shift operator L . Since $(D - a)(x^{i+1}e^{ax}) = x^i e^{ax}$, we see that $\text{Span}(e^{ax}, xe^{ax}, \dots, x^{k-1}e^{ax}) \subseteq \text{Ker}(D - a)^k$. But the functions $e^{ax}, xe^{ax}, \dots, x^{k-1}e^{ax}$ are linearly independent and so, since $\dim \text{Ker}(D - a)^k \leq k$, they are a basis for $\text{Ker}(D - a)^k$. Hence, for example, $\text{Ker}(D - 1)(D - 2)^2 \text{Ker}(D - 3)^3$ is 5-dimensional with basis $e^x, e^{2x}, xe^{2x}, e^{3x}, xe^{3x}, x^2e^{3x}$.

In the case of the left-shift operator L we have, in the case $a \neq 0$,

$$(L - a)(n^{i+1}a^n) \in \text{Span}((a^n), (na^n), \dots, (n^i a^n))$$

so that $\text{Span}((a^n), (na^n), \dots, (n^{k-1}a^n)) \subseteq \text{Ker}(L - a)^k$. But these sequences are linearly independent and so are a basis of $\text{Ker}(L - a)^k$ since $\dim \text{Ker}(L - a)^k \leq k$. Hence, for example,

$$\dim(\text{Ker}(L - 1)(L - 2)^2(L - 3)^3) = 5$$

with basis $(1), (2^n), (n2^n), (3^n), (n3^n), (n^2 3^n)$.

As another example, consider the problem of finding a formula for $s_n = 1^3 + 2^3 + \dots + n^3$. Let $s = (s_n)$. Then $(L - 1)(s) = ((n + 1)^3 - n^3) = 3(n^2) + 3(n) + (1)$ which is in the kernel of $(L - 1)^3$. Hence s is in the kernel of $(L - 1)^4$ so that there are constants A, B, C, D such that $s_n = A + Bn + Cn^2 + Dn^3$. The constants A, B, C, D can be found by solving the system of equations obtained by setting $n = 0, 1, 2, 3$.

Particular solutions to non-homogeneous difference and recurrence equations can be sometimes found by transforming the non-homogeneous equation into a homogeneous one. For example, the equation $(L - 1)(L - 2)x = (1)$ can be transformed into a homogeneous one by applying the operator $L - 1$ to both sides of the equation obtaining $(L - 1)^2(L - 2)x = 0$. The operator $L - 1$ was chosen to kill the right-hand side of the equation. Thus $x_n = a + bn + c2^n$. The term $a + c2^n$ can be omitted as it is in the kernel of $(L - 1)(L - 2)$. We thus look for a particular solution of the form $x_n = bn$. The constant b can be found by substitution in the given non-homogeneous equation. A similar procedure applies to differential equations.

2

Several authors have contributed bits and pieces to my understanding of this subject, but no author has given a clean linear presentation, so I will here attempt my own integration.

1. The Jordan Block.

Consider a Jordan Block of dimension 3×3 ,

$$J = \begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = C + \lambda_1 I$$

We have isolated a matrix called C here. As shown on page 33 of Matthews (apart from transpose), this is the companion matrix for the minimum polynomial $f(x) = x^3$. That is, we have that $C^3 = 0$. As you keep raising the power of C , you keep shifting the row of 1's to the upper right until it finally shifts completely away and you have a completely zero matrix left. Here is a more detailed example with $k = 4$

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad C^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad C^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad C^4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

To see why this is happening, note that you can write $C_{jk} = \delta_{j+1,k}$ to represent the initial diagonal of ones in the first off diagonal. Then we have

$$(C^2)_{jk} = \sum_m C_{jm} C_{mk} = \sum_m \delta_{j+1,m} \delta_{m+1,k} = \delta_{j+2,k}$$

$$(C^3)_{jk} = \sum_m C_{jm}^2 C_{mk} = \sum_m \delta_{j+2,m} \delta_{m+1,k} = \delta_{j+3,k} \quad \text{and so on}$$

Recall the way you are supposed to find the min poly of a matrix: keep raising the power until you get a linear combination of lower powers. Normally when you do this with some arbitrary matrix C , you get something like this at some point,

$$C^k = -a_0 C^0 - a_1 C^1 + \dots - a_{k-1} C^{k-1}$$

with some coefficients a_i . We can rewrite this as (think $a_k = 1$)

$$a_0 C^0 + a_1 C^1 + \dots + a_{k-1} C^{k-1} + C^k = 0$$

or

$$f(C) = 0 \quad \text{where } f(x) = x^k + a_{k-1}x^{k-1} + \dots + a_0$$

and then $f(x)$ is the min poly for matrix C . In our special case, we get this:

$$C^k = 0 \quad k = 4 \quad f(x) = x^k$$

Notice that the lower powers C, C^2, C^3 cannot be expressed as lin coms of previous powers, because the ones are in the wrong place. That is WHY the min poly is C^k .

So far, then, we have shown this fact:

Theorem 1: The min poly $m_C(\lambda)$ of the $k \times k$ matrix $(J - \lambda_1 I) = C$ is given by $m_C(\lambda) = \lambda^k$, and $m_C(C) = C^k = 0$.

Theorem 2: The min poly $m_J(\lambda)$ of the $k \times k$ matrix J is given by $(\lambda - \lambda_1)^k$.

Proof: We know that min poly of C is $m(C) = C^k$ which we can write as $(J - \lambda_1 I)^k$. We can then expand this thing to get $J^k - \lambda_1 J^{k-1} + \dots = 0$, and this then must be the min poly for the matrix J . Any lower power of J (below J^k) cannot be expressed as a lincom of lower powers. So $m_J(J) = (J - \lambda_1 I)^k$. But now write this in terms of a scalar argument λ to get $m_J(\lambda) = (\lambda - \lambda_1)^k$.

Theorem 3: The char poly $ch_J(x)$ of the $k \times k$ matrix J is given by $ch_J(\lambda) = (\lambda - \lambda_1)^k$, so we can regard k as the algebraic multiplicity of the eigenvalue λ_1 of the matrix J .

Proof: $ch_J(\lambda) = \det(J - \lambda I) = (\lambda - \lambda_1)^k$, just manually do the det, the off diagonal 1's have no effect.

Comment: The Jordan Block is constructed just to get this very result: the min poly = the char poly. Usually the min poly has a lower exponent, but not here!

Theorem 4: The matrix J has only one eigenvector $(x, 0, 0, 0, \dots)$ and thus has geometric multiplicity = 1.

Proof: The eigenvector problem is this: $Jv = \lambda_1 v$, which is the same as $Cv = 0$ which tells us that $v_{i+1} = 0$ for $i = 0, 1, 2, \dots$. Thus, our eigenvector is $(v_0, 0, 0, 0, \dots)$. We can regard this as the one eigenvector of the nullspace of matrix $C = (J - \lambda_1 I)$. The nullity of this space is then 1.

Conclusions reached concerning the Jordan block J of dimension k with diagonal elements λ_1 :

- (1) J has char poly = min poly = $(\lambda - \lambda_1)^k$
- (2) J has one eigenvalue λ_1 and that eigenvalue has an algebraic multiplicity of k
- (3) J has a geometric multiplicity of 1 and its only eigenvector is $(v_0, 0, 0, 0, \dots)$

2. Direct Sum of Jordan Blocks with same eigenvalue λ_1

Suppose we now construct a larger matrix ($\dim k = k_1 + k_2$) by doing a direct sum of 2 Jordan blocks, both of which have the SAME eigenvalue λ_1 :

$$A = J_{k_1} \oplus J_{k_2}$$

The first block has min poly $(\lambda - \lambda_1)^{k_1}$, the second has $(\lambda - \lambda_1)^{k_2}$. What is the min poly for the larger matrix? The result is derived on page 36 of Matthews. If we let $C = A \oplus B$, then he shows that

$$m_C = \text{lcm}(m_A, m_B) \quad \text{' lcm = least common multiple'}$$

We need to pause to prove this important result:

Theorem 5: If $C = A \oplus B$, then $m_C = \text{lcm}(m_A, m_B)$.

Proof: Assume that m_C is some polynomial $f(x)$. Then we know that:

$$f(C) = 0 = f(A \oplus B) = f(A) \oplus f(B) \text{ which } \Rightarrow f(A) = 0 \text{ and } f(B) = 0.$$

This says that $m_A \mid f$ and $m_B \mid f$. It says $m_A \mid f$, for example, because m_A is the lowest degree poly that satisfies $m_A(A) = 0$, so if $f(A)$ is some *other* poly with $f(A) = 0$, then m_A must divide into f . Now, we want to know the *lowest degree poly* $f(x)$ that can be divided by *both* $m_A(x)$ and $m_B(x)$. This is exactly what $\text{lcm}(m_A(x), m_B(x))$ is. And if the $f(x)$ so found is in fact the poly of lowest degree, then it must be the min poly of C . There can be no poly of lower degree that satisfies the requirement $m_A \mid f$ and $m_B \mid f$.

Now we apply this theorem to our direct product situation. We have $m_A = (\lambda - \lambda_1)^{k_1}$ and $m_B = (\lambda - \lambda_1)^{k_2}$. The answer is that $\text{lcm}(m_A, m_B) = (\lambda - \lambda_1)^{\text{MAX}(k_1, k_2)}$. For example, if we have $m_A = (\lambda - \lambda_1)^2$ and we have $m_B = (\lambda - \lambda_1)^3$, then $(\lambda - \lambda_1)^3$ can be divided by both m_A and m_B . So we can easily extend this idea to show the following

Theorem 6: Suppose we construct a matrix by combining multiple Jordan Blocks of various sizes, but which all have the same eigenvalue λ_1 ,

$$A = J_{k_1} \oplus J_{k_2} \oplus J_{k_3} \oplus \dots$$

Then the min poly for A is given by $m_A = (\lambda - \lambda_1)^{\text{MAX}(k_1, k_2, k_3, \dots)}$. This conclusion is reinforced on page 37 of Matthews.

Comment: If you are staring at the matrix A just described, you can easily see which submatrix has the largest dimension, call it $k_{\max} = \text{MAX}(k_1, k_2, \dots)$. Later, when we combine matrices A_i for different eigenvalues λ_i to make a still larger matrix B , and when we talk about the overall min poly for this larger matrix B , it will have a factor $(\lambda - \lambda_i)^{k_{\max}}$ for our particular submatrix A . So, the size of largest Jordan block in each λ manifold equals the exponent of the overall min poly for the matrix A .

Theorem 7: For the same matrix A shown in Theorem 6, the char poly is given by $(\lambda - \lambda_1)^{k_1 + k_2 + \dots}$.

Proof: When we have *any* block diagonal form, we know that we can multiply the sub-determinants to get the total determinant, so in this case we can multiply the char polys of the sub-matrices, and that then leads to a sum of exponents as shown.

3. Primary Decomposition

Suppose we are given some matrix T and we have computed its char poly and min poly to be:

$$\text{ch}_T(\lambda) = (\lambda - \lambda_1)^{a_1} (\lambda - \lambda_2)^{a_2}$$

$$m_T(\lambda) = (\lambda - \lambda_1)^{b_1} (\lambda - \lambda_2)^{b_2}$$

where we know that $a_i \geq b_i$ because we know that $m_T(\lambda) \mid \text{ch}_T(\lambda)$. Then:

Theorem 8: According to the Primary Decomposition Theorem on page 61 of Matthews, we can decompose our vector space V , in which this matrix T acts, into a pair of subspaces,

$$V = V_1 \oplus V_2$$

where

$$V_1 = N[(T - \lambda_1 I)^{b_1}] \quad \text{with } v = a_1$$

$$V_2 = N[(T - \lambda_2 I)^{b_2}] \quad \text{with } v = a_2$$

$N[X]$ means "Nullspace of matrix X "

where the exponent b_i is the $k_{\max}$ mentioned earlier for each subspace $i = 1$ and 2 . So b_i will be the size of the largest Jordan block in each of the two subspaces. Remember: b_i is the min poly exponent. The a_i are the algebraic multiplicities of each subspace.

Nothing is said here about geometric multiplicity, but we know that geo mult is the number of Jordan blocks in each subspace. We also know these facts:

$$N[(T - \lambda_1 I)^1] \text{ has } v = \text{geo mult } 1$$

$$N[(T - \lambda_2 I)^1] \text{ has } v = \text{geo mult } 2$$

These nullspaces are the "eigenmanifolds" of Stakgold. Here we have unity exponents.

4. The Cyclic Subspace Business

Following Matthews page 54, we imagine some "vector specific" min poly $m_{T,v}(T)v = 0$ rather than the more general min poly $m_T(T) = 0$. This specific thing has to divide the normal min poly.

Theorem 9: If f is any polynomial and v is some vector in V (the space in which T acts), then the set of all vectors of the form $f(T)v$ forms a *subspace* of V . This is called the **T -cyclic subspace generated by the vector v** , and this subspace is a **T -invariant subspace**, and it is represented as $C_{T,v}$.

Proof: I think the subspace part follows from usual linearity stuff. That is, $f_1(T)$ and $f_2(T)$ make a sum that is in the same space, so it is closed under addition, etc. Now suppose w is in $C_{T,v}$. Then $w = f(T)v$ for some polynomial f . Then $Tw = Tf(T)v = g(T)v$ which is also in the space since g is just some other polynomial.

Theorem 10: In the situation described in Theorem 9, assume that for vector v we have some vector-specific min poly $m_{T,v}(T)$ of degree k . Recall that v is the generator vector of the subspace $C_{T,v}$. Then the following vectors form a basis in $C_{T,v}$: $\{ v, Tv, T^2v, \dots, T^{k-1}v \}$ This is the **T-cyclic basis**.

This theorem has a proof on page 55,56 of Matthews which I have not read.

Theorem 11: Again in the situation of the previous two theorems, if we work in the T-cyclic basis, then the matrix T is none other than the companion matrix of the min poly $m_{T,v}$.

This is shown on page 56 Matthews.

Application of the Previous Theorems. Suppose $m_{T,v}(\lambda) = \lambda^k$. This is a poly of degree k , and we can form the T-cyclic basis of theorem 10: $C_{T,v} : \{ v, T^1v, T^2v, \dots, T^{k-1}v \}$ where we use T everywhere. Moreover, the matrix T in this cyclic basis is given (according to theorem 11) by the companion matrix which corresponds to $m_{T,v}(\lambda) = \lambda^k$, and we know from earlier work that this matrix is exactly the matrix C we discussed in section 1 above, the matrix with all zeros except for the first off-diagonal of ones.

Now if we define T as the shift $T' = (T - \lambda_1 I)$, we can do what we did in section 1 above and claim that $m_{T',v}(\lambda) = (\lambda - \lambda_1)^k$ is the v -specific min poly for operator T . Furthermore, the T-cyclic basis is now

$$C_{T,v} : \{ v, (T - \lambda_1 I)v, (T - \lambda_1 I)^2v, \dots, (T - \lambda_1 I)^{k-1}v \}$$

The matrix $T = T' + \lambda_1 I$ is then none other than J , the $k \times k$ Jordan Block matrix with eigenvalue λ_1 .

Comments: So, we have now learned that the Jordan Block arises when you take as your basis vectors the strange set $\{ v, (T - \lambda_1 I)v, (T - \lambda_1 I)^2v, \dots, (T - \lambda_1 I)^{k-1}v \}$.

Status: At this stopping point, we have "come down from above" with the primary decomposition to get a set of subspaces V_1 and V_2 , and we have "come up from below" with some strange cyclic subspace gizmo to create one Jordan block. It remains to connect these two worlds.

Status 1/7/05. Things are now pretty well connected, I think I am done with this subject for now, after spending quite a few days on it. Results are all in a binder and in a directory.

3

This is a 10 page set of lecture notes from www.stanford.edu/class/ee263/jcf.pdf from Prof. Steve Boyd, see site <http://www.stanford.edu/class/ee263/> for more stuff. This is a good place to go for lots of stuff. This was one of the first sets of notes I got for the Jordan stuff, and I did not understand them the first time, so now having done all the Matthews work, we can come back and look at what Boyd has to say.

Page 10-2 Here we get $J = T^{-1}AT$ and he shows the 1's of the J block on the upper off diagonal, whereas Matthews puts it on the lower. I am not much concerned about this transpose issue. He calls the Jordan block "upper bidiagonal". All comments on this page are now clear. When I first read it, I did not even know if the matrix J was the same dimension as matrix A. He refers to T as a similarity.

Why does he restrict to A having real matrix elements? I don't think this is necessary, but I agree that if you do, J can still be complex since eigenvalues can be complex.

Page 10-3. Comment that no one uses JCF for numerical work, only for exact work.

Then he writes the char poly for A with variable s instead of x or λ , and his n_i are my a_i , the char poly exponents which are of course the algmults.

Next, he comments that if all the $n_i = 1$, A can be diagonalized. I know this is true, A then has all eigenvalues different.

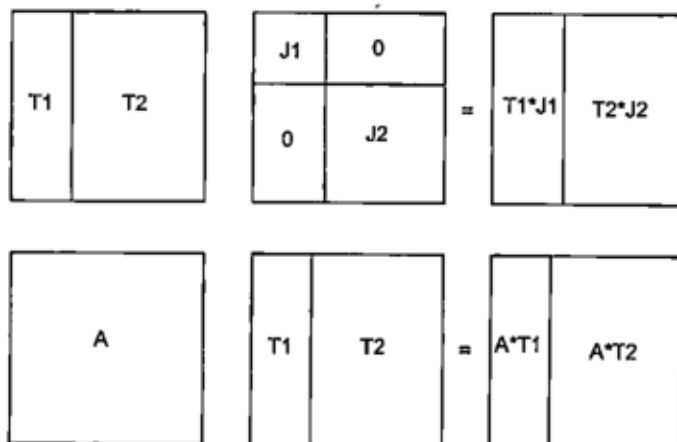
Next, $\dim \text{Ker } p = \# \text{ Jordan blocks}$. I now understand this to be true since each Jordan block has $\text{geomult} = 1$, and $\dim \text{Ker } p = \text{geomult of the entire A}$. What he really means is that $\dim \text{Ker } p_i$ is the number of Jordan blocks in the eigenmanifold $p_i^{a_i}$. If you then do this over all eigenmanifolds, then the total number of Jordan blocks in the final J is in fact $\sum_i \dim \text{Ker } p_i$. So we tick him for vagueness on this point.

Now comes his famous "more generally" claim. He is talking about $\dim \text{Ker } p^k$ here for $k = 1, 2, \dots$, and these are the $v(p^k)$ of Matthews. Consider our Matthew's second example which we just wrote up with the Figure 1 Visio picture. In that case we had $v(p^3) = 6$ and $n_1 = 6$ for our one eigenmanifold. So in this case, his equation is wrong. He says $\dim \text{Ker } p^3 = \min(3, 6) = 3$. Maybe on the left he means the sum over all such spaces, in which case his $\text{LHS} = \sum_i \dim \text{Ker } p_i^k = \sum_i v(p_i^k)$ but this is nothing useful to anyone! So I think our guy has screwed up here. He comments about the "sizes of the Jordan blocks". I know these are the heights of the columns in the dot diagram. Well, it is true that we have $v(p^k) = \dim \text{Ker } p^k$ and then from these numbers we can compute the $v_{k,p}$ numbers using the subtraction rule, and from those we can get the dot diagram, and from that we can get the sizes of the Jordan blocks. But we do not have the simple claim he makes that $v(p^k) = \text{some expression}$. I cannot even ascribe any meaning to his summation notation. But the general idea is that from the $v(p^k)$ you can figure out the sizes of the Jordan blocks, that claim is true.

Next, he claims that $T p_i(A) T^{-1} = (J - \lambda_i I)$ which I know is the companion matrix. He then shows that as you take higher powers of this matrix, it soon becomes all zeros. This is just the idea that the min poly of a C matrix (and of the J matrix) is the full dimension of the matrix. In his case, the min poly is of degree 3 and we can say $C^3 = 0$.

His last statement concerns an off diagonal power $(J_j - \lambda_i)^k$. I could verify this I suppose by brute force computation, but I don't see any use for the result, so I won't try to verify it.

Page 10-4 Now we are going to examine the matrix T which Matthews calls P. Instead of expressing T in columns, he puts it into groups of columns called T_i . Well, here is an interesting fact to keep in mind:



So if we are thinking about $A T = T [J_1 \oplus J_2]$, we can break this into two separate problems:

$$AT_1 = T_1 J_1 \quad \text{and} \quad AT_2 = T_2 J_2$$

so I agree with his general breakdown into $AT_i = T_i J_i$. Then you take one of these groups of columns and say this:

$$T_i = [v_{i1} \ v_{i2} \ \dots \ v_{ia_i}] \quad \text{where } a_i = \text{the algmult of the } J \text{ block } i.$$

These v's are then single column vectors. We then get his nice result that the first column (only) of one a T_i group is a true eigenvector. The other columns have a slightly different equation and they are called the generalized eigenvectors. I have derived all this stuff. Now in Matthews, we got a similar situation, but there the generalized eigenvectors formed the cyclic basis of a T cyclic subspace for J.

Well, I think I can make the connection as follows. His general equation is this:

$$(A - \lambda_i) v_{ij} = v_{i,j-1} \quad j > 1 \text{ up to } j = a \quad \Rightarrow p_i v_{ij} = v_{i,j-1}$$

so we have: $p_i v_{i2} = v_{i1}, \ p_i v_{i3} = v_{i2}, \ p_i v_{i4} = v_{i3} \ \dots \ p_i v_{ia} = v_{i,a-1}$

which I can rewrite as (for the case $a = 4$)

$$(p_i)^3 v_{i4} = (p_i)^2 v_{i3} = (p_i)^1 v_{i2} = v_{i1}$$

$$(p_i)^2 v_{i4} = (p_i) v_{i3} = v_{i2}$$

$$(p_i)^1 v_{i4} = v_{i3}$$

$$(p_i)^0 v_{i4} = v_{i4}$$

So here we see the Matthews idea that you get the basis vectors of the J_4 space by applying powers of p to a generator vector which here is v_{i4} , and the highest power $(p_i)^3 v_{i4}$ then gives you the true eigenvector

v_i 1, and all the other powers give you those generalized eigenvectors, so we are on the same page! What you would really like is to first find v_{14} , then generate the various other guys by applying powers of p .

Page 10-5 Now suddenly we are looking at a Linear Differential System LDS $dX/dt = AX$ where A is a matrix. This was also touched upon in Matthews. Here Boyd suggests doing a change of basis as shown to get a much simpler problem in which the matrix J is used. Let's look more into this:

$$dX/dt = AX \Rightarrow T^{-1}dX/dt = T^{-1}A T T^{-1}X \Rightarrow dX'/dt = [J_1 \oplus J_2]X'$$

We can now break our tall vector X' into $X' = [\dots\dots]^t = [x_1' \ x_2' \ \dots]^t$ and the problem then decouples into this set of smaller problems:

$$dx'_i/dt = J_i x'_i \quad // \text{ just as he says.}$$

Now, the i index tells you which Jordan block. Let's look just at a single Jordan block for some fixed i , and also let's get rid of the primes for now, knowing where we are. Then we have

$$dx/dt = J x$$

and if we assume Laplace $s = i\omega$ sine action, we can write this as

$$J x = s x$$

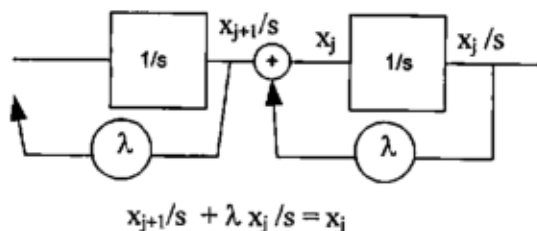
But remember that $J = C + \lambda I$ and that $C_{jk} = \delta_{j+1,k}$ so we can write our equation as

$$(C + \lambda I) x = s x \Rightarrow \sum_k (\delta_{j+1,k} + \lambda \delta_{j,k}) x_k = s x_j$$

$$\Rightarrow x_{j+1} + \lambda x_j = s x_j$$

$$\Rightarrow x_{j+1}/s + \lambda x_j/s = x_j$$

and we then get this **Jordan chain** picture



where we have one "stage" for each component of the vector x . Just think of the $1/s$ box as an integration operator if you like. So within each eigenmanifold, you can solve your system in this manner, then combine all the solutions into the full X' vector, then transform back to the original space. Lot's of work.

Page 10-6 Resolvent, Expo of J. Now we look at the meaning of p^{-1} . This is a matrix. It is the inverse of p , and we find a simple closed form result for p^{-1} . When we write $p^{-1} p = I$, we have product of two triangulars is a triangular. He defines F_i to mean a diagonal of 1's and this allows the nice form shown for p^{-1} .

He next shows that you can compute $\exp(tJ_\lambda)$, and his result is like Matthew's result on page 83, except we have the transpose issue in the definition of J .

What is the "resolvent"? It is exactly the inverse of p , something I am tempted to call a propagator. But resolvent is a standard name for the inverse secular operator.

Page 10-7 Generalized modes. We return to our system of first order ODE's as shown. We can solve it as shown on page bottom using the expo stuff (as Matthews did). If we pick the special $t=0$ starting vector as shown at the top of the page, then for all later time, the vector $x(t)$ is controlled only by one of the Jordan blocks of our matrix A . You think of this as if it were a vibrational normal mode. Then at the bottom you see that with a general starting vector $x(0)$, you have activity in all the "modes".

I can see that this is a whole new subject that one could study, I never heard of it before.

Page 10-8 Here he simple states the C-H theorem.

Page 10-9,10 The corollary here is that or arbitrary powers of A , you have a basis that is a finite number of powers, as determined by C-H (or perhaps by the min poly, but that is not mentioned). He directly proves this little corollary based on the C-H which makes $\chi(A) = 0$. If A is invertible, the same claim of finite basis applies to negative powers of A .

Page 10-11. His proof of C-H. He first assumes A can be diagonalized into Λ , then C-H is quite easy to show since all matrices are diagonal. Each one has a 0 somewhere on the diagonal and the effect of all the zeros is to kill the result so $\chi(\Lambda) = 0$, so that means $\chi(A) = 0$. More generally, we can get A into J form. Then you can show that $\chi(J_i) = 0$ for each i as shown, and then the direct sum business tells you that the full $\chi(J) = 0$, and so $\chi(A) = 0$. This is a nice proof I think! But you have to use the Jordan form, so it depends where you are starting your circle of logic.

Comments: This is a nice little lecture I think, lots of good topics are reached. He presents a direct method for computing the matrix T . You cannot say that he "proves" the Jordan Form in the sense that Matthews does. He assumes it, and then computes T with that assumption. He then touches on these applications: system of first order DE's, resolvent, generalized modes, and the C-H idea and what it implies.

Comments on the EE263 Lecture 10 Document

These are my original notes on this lecture when I knew very little about the subject.

These notes add a small amount to my picture which is still very incomplete. Page 2 states the Jordan block form which I now see better. At first I thought maybe the T were non-square, but not so. So page 2 is just fine.

Page 3 is full of mysteries. I agree with the fact that $\dim N(\lambda I - A) =$ geometric multiplicity of the manifold associated with eigenvalue λ . And I know from the above notes that this equals the number of Jordan blocks for this λ (but I don't know why that should be so). The next line where author says "more generally" -- this means nothing to me at all. I am not grokking a major point here, and author is not helping me at all with words.

Now I do know from the TL notes above that in $K(A)$ we have factors like $(\lambda I - A)^r$ where r is the algebraic multiplicity. And I know that there is a thing called "the minimal polynomial" which has the same factor but possibly raised to a lower power $(\lambda I - A)^s$. In the TL notes page 8 I agreed with the fact that this exponent s must be max of the sizes of all the Jordan blocks associated with this particular λ . But this does not seem to relate.

Recall from earlier work that $\text{Ker}(A - I\lambda)$ is the nullspace whose dimension m is the geomult of this eigenvalue λ , the number of different eigenvectors. We are somehow breaking this nullspace down into m distinct nullspaces, each of which has geomult = 1. This is exactly TL's Exercise 3. And we also know that the sum of the dims of these subspaces is r , the algmult. So we are doing a factorization like so:

$$(\lambda I - A)^6 = (\lambda I - A)^1 (\lambda I - A)^2 (\lambda I - A)^3 \quad 6 = \text{algmult} \quad 3 = \text{geomult}$$

and we associate each factor with one of these new nullspaces.

AHA...! (as one says). We know that the $K(A)$ equation is invariant under similarity, that is very easy to show and I have done it. So consider the above product in the language of J

$$(\lambda I - J)^6 = (\lambda I - J)^1 (\lambda I - J)^2 (\lambda I - J)^3 \quad 6 = \text{algmult} \quad 3 = \text{geomult}$$

Suppose for the moment that the entire J has only this one eigenvalue λ , but there are three Jordan blocks which we will call $J = \text{diag}(J_1, J_2, J_3)$ which are 1, 2 and 3 in dimension. In the above equation, I and J are both in block diagonal form, so this means we have a direct sum view and we can look only at a particular block, in which case we get this:

$$(\lambda I - J_i)^6 = (\lambda I - J_i)^1 (\lambda I - J_i)^2 (\lambda I - J_i)^3 \quad 6 = \text{algmult} \quad 3 = \text{geomult} \quad i = 1, 2 \text{ or } 3$$

Our char equation is just $(\lambda I - J_i)^6 = 0$ for our hypothetical matrix. But within each nullspace, we want to have $(\lambda I - J_i)^{\dim NS} = 0$. This will then say that $(\lambda I - J_i)^{\dim NS} x = 0$ for all x , so then

$$\dim \text{Ker} [(\lambda I - J_i)^{\dim NS}] = \dim NS \quad \text{write as} \quad \dim N(\lambda I - A)^k = k$$

Now our page 3 notes show that, with J_i defined as it is (the Jordan Block), we do in fact get $(\lambda I - J_i)^{\dim NS} = 0$. On each power, the ones slide up one diagonal, and then they slide off, just the way my little UH_n theorem says they should. So this at least motivates the form of the Jordan Block. On the right above, we have something that at least resembles what is on page 10-3 of the notes. Now we could never talk about $\dim NS$ being larger than the algmult of that λ .

Now in a more general matrix with several distinct λ_i , we do get the "off diagonal" powers as shown on the bottom of page 10-3. I have not checked the matrix on the right, but I can see it will be something that is non-zero.

Page 10-4

Now we get another small iota of insight. If we simply assume that some T exists to convert A to J , what can we say about this matrix T ? Well, we certainly can say $AT = TJ$, the way we used to say $A\Lambda = T\Lambda$ where Λ was a diagonal matrix. Because J is in block diagonal form, we can write this as $AT_i = T_i J_i$ where the J_i are the small square Jordan blocks, and the T_i are groups of column vectors of T which we can associate with that J_i . I have shown graphically how this works on that page. So we have a decoupling of the problem somewhat.

We then label the columns in each T_i . I like to say this as $T^{(i)} = [v_1^{(i)}, v_2^{(i)}, v_3^{(i)}, \dots, v_N^{(i)}]$. We need to maintain some kind of label "i" to say which T_i set of columns we are talking about. But then I leave off the i label and just think about ONE of these T blocks. We then find these interesting facts:

(1) The first column $v_1^{(i)}$ satisfies $A v_1^{(i)} = \lambda_i v_1^{(i)}$ and so it is a true eigenvector of λ_i . Remember that each Jordan block only has one true eigenvector, and this is it!!!

(2) The other columns satisfy equations like $A v_3^{(i)} = \lambda_i v_3^{(i)} + v_2^{(i)}$. This is *close* to an eigenvector equation, *but* we have the extra term on the right which is the previous column! We know that these other columns cannot be true eigenvectors since each block has only one, so these other columns are called the generalized eigenvectors. They are some kind of dummies to fill out the T block. It is not clear to me looking at this equation $A v_3^{(i)} = \lambda_i v_3^{(i)} + v_2^{(i)}$ whether or not there is a solution. I guess if there is not, that determines the width of the T block, which is to say, that determines the size of this Jordan block.

Page 10-8

I have to skip other pages because they are specialized to the application of this EE263 course at Stanford, but the last two pages 8 and 9 are good. He first makes a "statement" about what you mean by a poly with matrix coefficients, then he states the C-H theorem and gives a nice 2x2 example of it.

Page 10-9

The first part here is the "corollary" to C-H which says any power A^k can be written as a lincom of powers of A up to $N-1$, which I agree with. The second part is basically the same proof of "the corollary" that I put into my "matrix research" notes, but I think my details are much better, his is a little fudgy. Of course this proof depends on the C-H theorem.

Page 10-10.

This is a way to show that A^{-1} can also be expanded in the same lincom of powers of A , hence we can do any negative powers as claimed earlier if A is invertible.

Page 10-11

Now comes this guy's "proof" of the C-H theorem itself. It is all on the very bottom one line. It assumes that we can go to Jordan form, and then it uses the idea that the powers of the matrices vanish if you pick out the right one for each eigenvalue. Notice that the n_i here are NOT the algemults, they are the Jordan

block dimensions. So he uses *Jordan decomposition* to prove C-H, which is if nothing else interesting. Again, the fact that the secular equation is true in A or J is used.

I think I have now squeezed the most juice that I can out of each of these 10-page papers which I have printed and written on. But we need more info!

Lecture 10

Jordan canonical form

EE263

Autumn 2004

- Jordan canonical form
- generalized modes
- Cayley-Hamilton theorem

Jordan canonical form

what if A cannot be diagonalized?

any matrix $A \in \mathbf{R}^{n \times n}$ can be put in *Jordan canonical form* by a similarity transformation, i.e.

$$T^{-1}AT = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_q \end{bmatrix}$$

where

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbf{C}^{n_i \times n_i}$$

is called a *Jordan block* of size n_i with eigenvalue λ_i
(so $n = \sum_{i=1}^q n_i$) ✓

Big
Theorem

- J is upper bidiagonal ✓
- J diagonal is the special case of n Jordan blocks of size $n_i = 1$ ✓
- Jordan form is unique (up to permutations of the blocks) ✓
- can have multiple blocks with same eigenvalue
not necessarily of the same size!

note: JCF is a conceptual tool, never used in numerical computations!

$\chi(s) = \det(sI - A) = (s - \lambda_1)^{n_1} \dots (s - \lambda_q)^{n_q} \leftarrow$ algebraic multiplicities
 hence distinct eigenvalues $\Rightarrow n_i = 1 \Rightarrow A$ diagonalizable ✓
 $\downarrow$ nullspace
 $\dim \mathcal{N}(\lambda I - A)$ is the number of Jordan blocks with eigenvalue λ = the geometric multiplicity of λ .

more generally,

$$\dim \mathcal{N}(\lambda I - A)^k = \sum_{\lambda_i = \lambda} \min\{k, n_i\}$$

so from $\dim \mathcal{N}(\lambda I - A)^k$ for $k = 1, 2, \dots$ we can determine the sizes of the Jordan blocks associated with λ

• factor out T and T^{-1} , $\lambda I - A = T(\lambda I - J)T^{-1}$

• for, say, a block of size 3: Think $LH_1 \cdot LH_1 = LH_2$ ✓

$$\lambda_i I - J_i = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (\lambda_i I - J_i)^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\lambda_i I - J_i)^3 = 0$$

• for other blocks (say, size 3)

$$(\lambda_i I - J_j)^k = \begin{bmatrix} (\lambda_i - \lambda_j)^k & -k(\lambda_i - \lambda_j)^{k-1} & -k(k-1)/2(\lambda_i - \lambda_j)^{k-2} \\ 0 & (\lambda_j - \lambda_i)^k & -k(\lambda_j - \lambda_i)^{k-1} \\ 0 & 0 & (\lambda_j - \lambda_i)^k \end{bmatrix}$$

$$T(A \oplus B) = TA \oplus TB$$

suppose $T^{-1}AT = J = \text{diag}(J_1, \dots, J_q)$ ✓

$$[T] \left[\begin{array}{c|c|c} J_1 & & \\ \hline & J_2 & \\ \hline & & J_q \end{array} \right] = \left[\begin{array}{c|c|c} T_1 & & \\ \hline & T_2 & \\ \hline & & T_q \end{array} \right]$$

express T as

$$T = [T_1 \ T_2 \ \dots \ T_q] \checkmark$$

where $T_i \in \mathbb{C}^{n \times n_i}$ are the ^{for} columns of T associated with i th Jordan block J_i

we have $AT_i = T_i J_i$ ✓

$$[A] \underbrace{\begin{bmatrix} T_1 & T_2 & T_3 \end{bmatrix}}_T = \begin{bmatrix} T_1 J_1 & T_2 J_2 & T_3 J_3 \end{bmatrix}$$

let $T_i = [v_{i1} \ v_{i2} \ \dots \ v_{in_i}] \checkmark$
 $\underbrace{\hspace{10em}}_{\text{the columns}}$

$$\begin{bmatrix} a & a' \\ b & b' \\ c & c' \\ d & d' \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} \lambda a & a' + \lambda a'' \\ \lambda b & b' + \lambda b'' \\ \lambda c & c' + \lambda c'' \\ \vdots & \vdots \end{bmatrix}$$

$\underbrace{\hspace{10em}}_{T_2} \quad \quad \quad \underbrace{\hspace{10em}}_{= \lambda \cdot \vec{v}_1}$

then we have:

$$Av_{i1} = \lambda_i v_{i1}, \checkmark$$

i.e., the first column of each T_i is an eigenvector associated with e.v. λ_i ✓

for $j = 2, \dots, n_i$,

$$\checkmark Av_{ij} = v_{i,j-1} + \lambda_i v_{ij}$$

$$\begin{bmatrix} \lambda a & a + \lambda a' & a' + \lambda a'' \\ \lambda b & b + \lambda b' & b' + \lambda b'' \\ \lambda c & c + \lambda c' & c' + \lambda c'' \\ \lambda d & \vdots & \vdots \end{bmatrix}$$

$$\begin{cases} A \vec{v}_j = \vec{v}_{j-1} + \lambda_i \vec{v}_j \\ A \vec{v}_1 = \lambda_i \vec{v}_1 \end{cases} \quad \uparrow_{j>1}$$

the vectors $v_{i1}, \dots, v_{in_i}$ are sometimes called generalized eigenvectors

$$\text{eg, } A \vec{v}_2 = \lambda \vec{v}_2 + \vec{v}_1$$

$$A \vec{v}_3 = \lambda \vec{v}_3 + \vec{v}_2$$

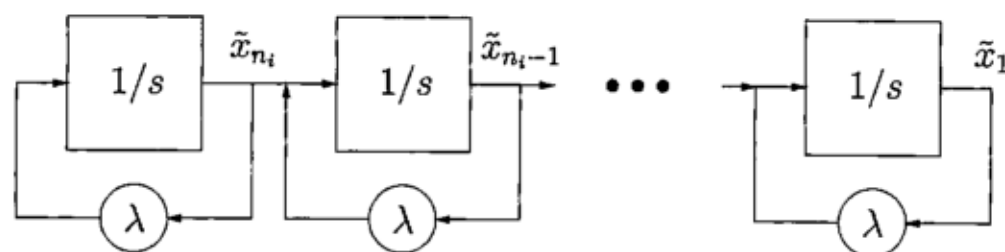
etc.

all OK

consider LDS $\dot{x} = Ax$

by change of coordinates $x = T\tilde{x}$, can put into form
 $\dot{\tilde{x}} = J\tilde{x}$

system is decomposed into independent 'Jordan block systems' $\dot{\tilde{x}}_i = J_i \tilde{x}_i$



Jordan blocks are sometimes called Jordan chains
 (block diagram shows why)



Resolvent, exponential of Jordan block

resolvent of $k \times k$ Jordan block with eigenvalue λ :

$$\begin{aligned}
 (sI - J_\lambda)^{-1} &= \begin{bmatrix} s - \lambda & -1 & & \\ & s - \lambda & \cdots & \\ & & \ddots & -1 \\ & & & s - \lambda \end{bmatrix}^{-1} \\
 &\stackrel{\text{resolvent ("propagator")}}{=} \begin{bmatrix} (s - \lambda)^{-1} & (s - \lambda)^{-2} & \cdots & (s - \lambda)^{-k} \\ & (s - \lambda)^{-1} & \cdots & (s - \lambda)^{-k+1} \\ & & \ddots & \vdots \\ & & & (s - \lambda)^{-1} \end{bmatrix} \\
 &= (s - \lambda)^{-1}I + (s - \lambda)^{-2}F_1 + \cdots + (s - \lambda)^{-k}F_{k-1}
 \end{aligned}$$

where F_i is the matrix with ones on the i th upper diagonal

by inverse Laplace transform, exponential is:

$$\begin{aligned}
 e^{J_\lambda t} &= e^{\lambda t} \left(I + tF_1 + \cdots + (t^{k-1}/(k-1)!)F_{k-1} \right) \\
 &= e^{\lambda t} \begin{bmatrix} 1 & t & \cdots & t^{k-1}/(k-1)! \\ & 1 & \cdots & t^{k-2}/(k-2)! \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}
 \end{aligned}$$

Jordan blocks yield:

- repeated poles in resolvent ✓
- terms of form $t^p e^{\lambda t}$ in e^{At} ✓

OK.

Generalized modes

consider $\dot{x} = Ax$, with

$$x(0) = a_1 v_{i1} + \cdots + a_{n_i} v_{in_i} = T_i a$$

then $x(t) = \checkmark T e^{Jt} \tilde{x}(0) = T_i e^{J_i t} a$

- trajectory stays in span of generalized eigenvectors
- coefficients have form $p(t)e^{\lambda t}$, where p is polynomial
- such solutions are called *generalized modes* of the system

with general $x(0)$ we can write

$$x(t) = e^{At} x(0) = \checkmark (T e^{Jt} T^{-1}) \checkmark x(0) = \sum_{i=1}^q T_i e^{J_i t} (S_i^T x(0))$$

where

$$T^{-1} = \begin{bmatrix} S_1^T \\ \vdots \\ S_q^T \end{bmatrix}$$

hence: all solutions of $\dot{x} = Ax$ are linear combinations of (generalized) modes

$$A^5 = 3A^4 - 2A^3 -$$

$$\begin{aligned} A^6 &= (A^4 - \sim -)A \\ &= (A^5 + A^{\sim} - -) \\ &= (\sim -) \end{aligned}$$

$$A^{900} = A^{9 \cdot N + \sim} =$$

$$\begin{aligned} A^{100} &= A^{95} A^5 = A^{95} (A^4 - - - -) \\ &= A^{94} (A^5 \sim -) \end{aligned}$$

Cayley-Hamilton theorem

if $p(s) = a_0 + a_1s + \cdots + a_k s^k$ is a polynomial and $A \in \mathbf{R}^{n \times n}$, we define $p(A) = a_0I + a_1A + \cdots + a_k A^k$

Cayley-Hamilton theorem: for any $A \in \mathbf{R}^{n \times n}$ we have $\mathcal{X}(A) = 0$, where $\mathcal{X}(s) = \det(sI - A)$

example: with $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ we have

$\mathcal{X}(s) = s^2 - 5s - 2$, so

$$\begin{aligned}\mathcal{X}(A) &= A^2 - 5A - 2I \\ &= \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} - 5 \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 2I \\ &= 0\end{aligned}$$

corollary: for every $p \in \mathbf{Z}_+$, we have

$$A^p \in \text{span} \{ I, A, A^2, \dots, A^{n-1} \}$$

(and if A is invertible, also for $p \in \mathbf{Z}$)

i.e., every power of A can be expressed as linear combination of $I, A, \dots, A^{n-1}$

proof: divide $\mathcal{X}(s)$ into s^p to get $s^p = q(s)\mathcal{X}(s) + r(s)$
 $r = \alpha_0 + \alpha_1 s + \dots + \alpha_{n-1} s^{n-1}$ is remainder polynomial
 then

$$\begin{aligned} A^p &= q(A)\mathcal{X}(A) + r(A) \\ &= r(A) \\ &= \alpha_0 I + \alpha_1 A + \dots + \alpha_{n-1} A^{n-1} \end{aligned}$$

for $p = -1$: rewrite C-H theorem

$$\chi(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_0I = 0$$

as

$$I = A \left(-(a_1/a_0) - (a_2/a_0)A - \cdots - (1/a_0)A^{n-1} \right)$$

(A is invertible $\Leftrightarrow a_0 \neq 0$) so

$$A^{-1} = -(a_1/a_0)I - (a_2/a_0)A - \cdots - (1/a_0)A^{n-1}$$

i.e., inverse is linear combination of A^k , $k = 0, \dots, n-1$

Proof of C-H theorem

first assume A is diagonalizable: $T^{-1}AT = \Lambda$

$$\mathcal{X}(s) = (s - \lambda_1) \cdots (s - \lambda_n)$$

since

$$\mathcal{X}(A) = \mathcal{X}(T\Lambda T^{-1}) = T\mathcal{X}(\Lambda)T^{-1}$$

it suffices to show $\mathcal{X}(\Lambda) = 0$

$$\begin{aligned} \mathcal{X}(\Lambda) &= (\Lambda - \lambda_1 I) \cdots (\Lambda - \lambda_n I) \\ &= \text{diag}(0, \lambda_2 - \lambda_1, \dots, \lambda_n - \lambda_1) \cdots \\ &\quad \cdots \text{diag}(\lambda_1 - \lambda_n, \dots, \lambda_{n-1} - \lambda_n, 0) = 0 \end{aligned}$$

now let's do general case: $T^{-1}AT = J$

$$\mathcal{X}(s) = (s - \lambda_1)^{n_1} \cdots (s - \lambda_q)^{n_q}$$

suffices to show $\mathcal{X}(J_i) = 0$

$$\mathcal{X}(J_i) = (J_i - \lambda_1 I)^{n_1} \cdots \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ & & \ddots & \ddots \end{bmatrix}}_{(J_i - \lambda_i I)^{n_i}} \cdots (J_i - \lambda_q I)^{n_q} = 0$$

Now these notes are finally comprehensible to me. At first I did not know what was going on. Notation is all text, so a little ugly. The nice part is that it uses Mathematica to quickly answer questions. In this thing, the author directly computes the Jordan form for two sample matrices.

Here is how he does it:

(1) write your matrix A, find out its eigenvalues. You want to pick an A that has some repeated eigenvalues to make it non-trivial. He does this.

(2) compute the eigenvectors. Here he finds out that one double λ has only one eigenvector, the second is reported as all zeros. What does Maple do in this situation, let's see!

`[1, 2, {vector([1, 0, 0])}, [2, 1, {vector([0, 1, 1])}]]` value, mult, vector

So Maple just says only one eigenvector for $\lambda = 1$.

(3) next, you have to create the missing generalized eigenvector using

$$A \mathbf{v}_2^{(i)} = \lambda_i \mathbf{v}_2^{(i)} + \mathbf{v}_1^{(i)}$$

But, just let's now remove the (i) superscript and just think of the $\mathbf{v}_i$ as columns of the overall T. Then we have

$$(A - \lambda I) \mathbf{v}_2 = \mathbf{v}_1$$

and we know $\mathbf{v}_1 = (1, 0, 0)$ and $\lambda = 1$ and we know A. So we just tell Maple to solve this inhomogeneous equation for $\mathbf{v}_2$, and Maple does this and gives us $(x, 2, 1)$ for any x, but Mathematic says $(0, 2, 1)$ as a viable solution.

(4) You are now done. You have your T, and do similarity on A to get J. I suspect you need to always put 0 in when you can for an unspecified number as in the above case.

Finding the Jordan Canonical Form of a Matrix

Let's begin with a matrix A.

In[8]:=

```
Clear[JA];
A={1,1,-1},{0,0,2},{0,-1,3};
MatrixForm[A]
```

Out[8]=

```
1    1    -1
0    0    2
0    -1   3
```

We will put A in Jordan form. To do this we need to find a nonsingular matrix S such that $\text{Inv}(S)AS=J$, where J is the Jordan form. This matrix S will be made up of eigenvectors and "generalized eigenvectors" as we shall see.

First, we must find the eigenvalues of A.

In[9]:=

```
Eigenvalues[A]
```

Out[9]=

```
{1, 1, 2}
```

You notice that A has a repeated eigenvalue. This does not necessarily mean that A does not have a full complement of eigenvectors. Let's see.

In[10]:=

```
Eigenvectors[A]
```

Out[10]=

```
{{1, 0, 0}, {0, 0, 0}, {0, 1, 1}}
```

Notice that by listing an eigenvector of $\{0,0,0\}$, *Mathematica* is telling you that A does not have a full complement of eigenvectors. We must search for "generalized eigenvectors".

An eigenvector v solves the matrix equation $Av=\lambda v$ [or $(A-\lambda I)v=0$, where I is the identity matrix]. A generalized eigenvector $v(i)$ will solve the matrix equation

$$Av(i)=\lambda v(i) + v(i-1)$$

[note that $v(i)$ represents a subscript].

Here $v(i)$ is the generalized eigenvector corresponding to the eigenvector $v(i-1)$

This equation can be written $(A - \lambda(i) I)v(i) = v(i-1)$.

$v(1)$ =the first eigenvector, $v(2)$ =generalized eigenvector to go with $v(1)$, and $v(3)$ is the second eigenvector. A generalized eigenvector will go in each slot that *Mathematica* gives the zero vector as an eigenvector.

In[11]:=

```
LinearSolve[A-1 IdentityMatrix[3],{1,0,0}]
```

Out[11]=

```
{0, 2, 1}
```

We have just found $v(2)$, the generalized eigenvector to go with $v(1)=\{1,0,0\}$. We now have three vectors to form the nonsingular matrix S . $\text{Inv}(S)AS$ should give us the Jordan form of A .

In[12]:=

```
S=Transpose[{ {1,0,0}, {0,2,1}, {0,1,1} }];  
MatrixForm[S]
```

Out[12]=

```
1  0  0  
0  2  1  
0  1  1
```

In[13]:=

```
JA=Inverse[S].A.S;  
MatrixForm[JA]
```

Out[13]=

```
1  1  0  
0  1  0  
0  0  2
```

Let's try another one.

In[14]:=

```
Clear[B,S]
```

In[15]:=

```
B={ {2,-1,2,0}, {0,3,-1,0}, {0,1,1,0}, {0,1,-3,5} };  
MatrixForm[B]
```

Out[15]=

```

2    -1   2    0
0     3   -1   0
0     1    1    0
0     1   -3    5

```

`In[16]:=`

```
Eigenvalues[B]
```

`Out[16]=`

```
{2, 2, 2, 5}
```

You notice that B has a repeated eigenvalue. This does not necessarily mean that B does not have a full complement of eigenvectors. Let's see.

`In[17]:=`

```
Eigenvectors[B]
```

`Out[17]=`

```
{{1, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 1}}
```

Notice that by listing an eigenvector of {0,0,0,0} twice, *Mathematica* is telling you that B does not have a full complement of eigenvectors. We must search for 2 "generalized eigenvectors".

`In[18]:=`

```
LinearSolve[B-2 IdentityMatrix[4],{1,0,0,0}]
```

`Out[18]=`

```
{0, 1, 1, 2/3}
```

Let's keep going....

`In[19]:=`

```
LinearSolve[B-2 IdentityMatrix[4],{0,1,1,2/3}]
```

`Out[19]=`

```
{0, 2, 1, 5/9}
```

We now have our four vectors for the matrix S.

`In[20]:=`

```
S=Transpose[{{1,0,0,0},{0,1,1,2/3},{0,2,1,5/9},{0,0,0,1}}];
```

`MatrixForm[S]`

Out[20]=

1 0 0 0

0 1 2 0

0 1 1 0

2 5

- -

0 3 9 1

In[21]:=

`JB=Inverse[S].B.S;`

`MatrixForm[JB]`

Out[21]=

2 1 0 0

0 2 1 0

0 0 2 0

0 0 0 5

Up to Jordan Canonical Form

These are excellent notes on the subject of Minimal Polynomials and Jordan Canonical Form.

Def: an **annihilating polynomial** $p(\alpha) = 0$. In this section, think of $\alpha = A$ a square matrix. So this polynomial is like the $K(A) = 0$ result of the Cayley-Hamilton theorem.

Def: a **minimal polynomial** $m(\alpha)$ is the *lowest degree poly* that does $m(\alpha) = 0$. Clearly, $m(\alpha) \mid p(\alpha)$, which notation means that m divides evenly into p , no remainder. I have then stated Prop 2.

Prop 1: every matrix α has exactly one minimal polynomial. Proof is sketchy, seems reasonable.

On page 2, he writes the $K(\lambda)$ but his notation is $\chi_\alpha(t)$, which is of course $\det(\alpha - tI)$, secular. The r 's you see here are the algebraic multiplicities. The claim is that the minimal polynomial looks similar, but each exponent is something in the range (1, alg mult), or as he writes it $1 \leq s_i \leq r_i$.

I am unsure how you would show that $m(\alpha)$ really is an annihilator. You have to build it and see if it is zero or not. That is, is the matrix $m(\alpha)$ equal to the zero matrix? It certainly seems reasonable that such a thing might exist that is just a piece of $K(\alpha)$. [I now know how to find $m(\alpha)$ for a matrix α , so this is not as confusing as on first reading.]

If matrix is diagonal, then $m(\alpha) = K(\alpha)$. This is because all powers are 1 and you cannot go any lower.

Lemma 4 is Sylvester's triangle law of nullity, and I now see that **Ker** is the way some people say Nullspace. He actually proves this law right in situ, but I ignore the proof.

How does he use this Lemma on the bottom of page 3? There are three claims made at the bottom

(1) the first line says $\dim(V) = \dim \text{Nullspace of } (K(A))$. We know that $K(A) = 0$ from C-H, so we know that $K(A)x = 0$ for all x in V , so that certainly means that the dim of this nullspace is V , so OK.

(2) Here he is saying that $\dim(V)$ is less than the sum of the dims of the nullspaces of each factor, and this is just Sylvester generalized to many factors, so this is fine too. Think of the RHS as being $\dim N_1 + \dim N_2 + \text{etc} =$ the sum of dims of a bunch of nullspaces. The idea here is that $N(ABC\dots)$ cannot exceed the sum of the dimensions of the individual nullspaces.

(3) In the last line, you can just re-express this as $\dim(N_1 \oplus N_2 \dots)$ where you make a new space $N = N_1 \oplus N_2 \dots$ by just superposing the other nullspaces. So $\dim(V) \leq \text{RHS of last line}$. I think in the case that each null space has dimension 1, which is the case for diagonalizable A , then the $\leq$ is really $=$ because we are adding up N numbers that are all 1.

Still I am confused as to what the point is here. It is the sentence stated at the top, but I lose it.

On page 4 we learn this his word for "similarity" is "conjugates", this is the first time I have seen this usage. M&M called them similarity of collinearity. We are reminded here that nearly everything is conserved under a similarity, and it is just like a change of basis.

On page 5 we get the definition of a Jordan block and the idea that we are going to write A in block diagonal form which is a direct sum (I think I conjectured this direct sum notation early on when first reading M&M, and here it is in the flesh).

Page 6 then states the Big Theorem without proof:

Big Theorem: Any matrix A can be converted by a similarity to Jordan Canonical Form. In this form, you have the direct sum of blocks. Here are the key results:

- (1) the eigenvalues are on the diagonal and each eigenvalue appears a number of times equal to its algebraic multiplicity. Thus must be the case since $\det(J - \lambda I) = K(\lambda)$ = the usual thing.
- (2) A degenerate eigenvalue λ_1 can have several Jordan blocks of various sizes. The size of the largest block in this group tells you the exponent to use for the factor $(\lambda - \lambda_1)^s$ in the *minimal* polynomial.
- (3) The number of Jordan blocks for a degenerate eigenvalue λ_1 is the geometric multiplicity of that λ_1 .

On page 9 we have a 6x6 example, and he goes through all the above items for this example.

Warning is issued: even if two matrices how all invariants the same, they may not be similarity conjugates.

Now back to page 6-7.

- (1) This Jordan thing is MUCH stronger than the Schur theorem which just says you can transform into upper triangular with some similarity.
- (2) The block structure is akin to the notion of breaking a number into powers of primes.

Exercise 1: Here we look at just one Jordan block that is $d \times d$ for some λ . The key point I learn here is that this matrix only has one eigenvector! So it has geomult = 1 by its construction. That eigenvector is just $(1, 0, 0, \dots)$.

Exercise 2: Here we look at the direct sum of just two blocks C and D . Most things are obvious, but item b is where we have something new. You want to construct the minimal polynomial for $C \oplus D$. Suppose these two matrix C and D have the same eigenvalue. Suppose $K_C(\lambda)$ contains $(\lambda - \lambda_1)^3$ and $K_D(\lambda)$ has $(\lambda - \lambda_1)^2$. Well, we know the overall $K(\lambda)$ will have $(\lambda - \lambda_1)^5$, but what about the min pol $m(\lambda)$? Each of the above factors has to divide evenly into $m(\lambda)$ otherwise you can show that m is not a min pol. This is why we get the notion that $m = (\lambda - \lambda_1)^3$ because we can do both $(\lambda - \lambda_1)^3 / (\lambda - \lambda_1)^3$ and $(\lambda - \lambda_1)^3 / (\lambda - \lambda_1)^2$ with no remainder, so we have to take the MAX of the two exponents.

Exercise 3: Now we look at an arbitrary set of blocks all for the same λ . We know that each block has $\text{geo} = 1$, so if there are g blocks, then $\text{geo} = g$. So boom, we see why the geomult is the number of blocks in the big Jordan form result. Secondly, the power in the min poly is again going to be the max of all the powers of $(\lambda - \lambda_i)$ for each block to get that common divisor effect as above. Of course now $K(\lambda)$ has lots of factors, and this comment applies only to the factor for λ_i and its Jordan blocks.

Exercise 4: Now we go out to the full matrix and state the final results. Results are the same, they just apply separately for each λ_i manifold.

My Comments:

- (1) It would be nice to see a proof of the big theorem.
- (2) I don't understand why the minimal polynomial is important.

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 $\alpha : V \rightarrow V$ 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. ✓ OK.

Uniqueness: Suppose that m and m' are different minimal 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 . ✓

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 , $p(\alpha) = 0 \Leftrightarrow m_\alpha | p$.

m divides p with no remainder $\Rightarrow m$ is a factor of 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 $m_\alpha | p$. $\checkmark$ $\square$

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 $m_\alpha | \chi_\alpha$. $\checkmark$

κ 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

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

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

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*

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

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 $s_i \leq r_i$. $\checkmark$

$\checkmark$ ie, $m_\alpha | \chi_\alpha$

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 $i \in \{1, \dots, k\}$ 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. $\checkmark$

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

$$\checkmark \quad \dim \text{Ker}(\gamma \circ \beta) \leq \dim \text{Ker}(\gamma) + \dim \text{Ker}(\beta).$$

Kernel = Nullspace
← Sylvester's Lon, M13

Proof Observe that $\text{Ker}(\gamma \circ \beta) = \beta^{-1}(\text{Ker}(\gamma))$. (By using ' β^{-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

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

Applying the rank-nullity formula we get

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

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

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

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

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

→ he means related by similarity

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

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

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

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

The determinant of α is $\det(\alpha - 0.I) = \chi_\alpha(0)$. ✓ The rank is n minus the nullity; ✓ and the nullity is $\dim \text{Ker}(\alpha - 0.I)$, 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 $A_1, \dots, A_m$ be square matrices, and let

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

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

$$A_1 \oplus \cdots \oplus A_m. \quad \checkmark$$

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

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

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 $m, n_1, \dots, n_m \geq 1$ and complex numbers $\mu_1, \dots, \mu_m$ such that A is conjugate to

$$J_{\mu_1}^{(n_1)} \oplus \dots \oplus J_{\mu_m}^{(n_m)}. \quad (\dagger)$$

Moreover, if A is also conjugate to

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

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 $(\dagger)$ 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

Schur Form.

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. ~~OK~~.

✓ **Like prime factorization** Any natural number $n \geq 1$ is equal to $p_1 p_2 \cdots p_m$ for some sequence $p_1, \dots, p_m$ 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: ~~OK~~

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 $K(\lambda) = (\lambda - \lambda)^d$
- ✓ b. minimal polynomial $(t - \lambda)^d$ — why?
- ✓ c. characteristic polynomial $(t - \lambda)^d$, and that
- ✓ d. the geometric multiplicity of λ is 1. ✓

$$\begin{pmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$\begin{pmatrix} v_2 \\ v_3 \\ v_4 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} v_1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$
 only evec.

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 ,

$$s'' = \max\{s, s'\} \quad \checkmark \quad \text{eg, } s=2, s'=3$$

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 $\chi_{C \oplus D}(t) = \chi_C(t) \cdot \chi_D(t)$. 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, $g = \dim \text{Ker}(C - \lambda I)$. The same convention applies to g' and g'' .)

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

- a. the only eigenvalue is λ ✓
- ✓ b. the power of $(t - \lambda)$ in m_A is $\max\{d_1, \dots, d_g\}$
- ✓ c. the power of $(t - \lambda)$ in χ_A is $d_1 + \dots + d_g$
- ✓ d. the geometric multiplicity of λ is g .

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

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

Show that for the matrix A ,

- a. the eigenvalues are $\lambda_1, \dots, \lambda_k$
- b. the power of $(t - \lambda_i)$ in m_A is $\max\{d_i^1, \dots, d_i^{g_i}\}$
- c. the power of $(t - \lambda_i)$ in χ_A is $d_i^1 + \dots + d_i^{g_i}$
- 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 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

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

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. ✓

Gumau's Proof of Jordan Decomposition

This guy is Guido Mauas, a Student of Mathematics, Buenos Aires, Argentina. He "owns" the little chunk of the PlanetMath "encyclopedia". I now regard this just as a summary of the detailed work that Matthews has done.

Guido uses α_i as the b_i , the exponents of the min poly. The primary decomp result is then clearly stated. He then talks about a "restriction" of operator T to $\ker p^b$, just the way that Matthew's did (and that I do in my later document on this subject with Visio drawing).

He then looks at $p: \ker p^b \rightarrow \ker p^b$, which is a little different from what I have done, but OK. We look for a basis for $\ker p^b$, the space which has $\text{algmult} = a$. The basis of $\ker p^b$ is written as the union of the bases of the Jordan cyclic subspaces. Each of these is given as $B_{s,i}$ where i = overall eigenmanifold index, and s = the secondary cyclic subspace label. He writes $B_{s,i}$ as powers of p acting on some assumed generating vector which he calls $v_{s,i}$. The union of these is called just B_i .

Now finally we come to the last section of this 1 pager proof. Author wants to show directly that, given the cyclic basis above, we can derive that the Jordan Block looks like. I will do this in my own language:

(1) Define $p = (T - \lambda I)$ as usual. Write $p^{j+1} = p p^j = (T - \lambda)p^j$. Thus, $T p^j = p^{j+1} + \lambda p^j$.

(2) We now use this result to evaluate our matrix elements:

$$\begin{aligned} J_{kj} &= \langle p^k v | T | p^j v \rangle = \langle p^k v | T p^j v \rangle = \langle p^k v | (p^{j+1} + \lambda p^j) v \rangle = \delta_{k,j+1} + \lambda \delta_{k,j} \\ &= C_{jk} + \lambda I_{kj} = (C^T)_{kj} + \lambda (I)_{kj} \text{ which says } J = C^T + \lambda I. \end{aligned}$$

where I am referring back to my PhL notes on the companion C matrix. The only difference is that we have C instead of C^T . So basically our author has derived the shape of the J block.

Question : I have had to assume that $\langle p^k v | p^j v \rangle = \delta_{k,j}$. How can I verify this?

In a cyclic subspace as generated by vector v we have this basis: $\{ v, pv, p^2v, p^3v \dots \}$. This subject is discussed (thank goodness) on page 55 of Matthews.

Well, this is quite interesting. In the entire Matthews presentation, we deal only with vector spaces and the associated notions of spanning and linear independence. There is no norm, no metric, no scalar product! He has no need for such things. On page 56 when we want to know what the matrix looks like for our operator T , we find the columns of the matrix simply by applying the operator to the vectors of the basis. In this manner, he finds the companion matrix C that matches the min pol of vector v .

So let's go back to our derivation above to the point where we have:

$$T p^j v = p^{j+1} v + \lambda p^j v. \quad p = (T - \lambda I)$$

We now find the columns of our matrix T by applying T to the basis vectors one at a time. We make this *association* between basis vectors and column vectors:

$$\begin{aligned} p^0 v &\rightarrow [1, 0, 0, 0]^t \\ p^1 v &\rightarrow [0, 1, 0, 0]^t \end{aligned}$$

$$\text{Then } T(p^0 v) = p^1 v + \lambda p^0 v = [0, 1, 0, 0]^t + [\lambda, 0, 0, 0]^t = [\lambda, 1, 0, 0]^t$$

Now once we make the above association, then yes, we can say that under the scalar product induced by the associated E^n , our basis vectors are orthonormal, and then we do have $\langle p^k v | p^j v \rangle_{\text{IND}} = \delta_{kj}$. But this is not the obvious scalar product that you would get directly from E^n

$$\langle p^k v | p^j v \rangle_{\text{IND}} \neq [p^k v]^T [p^j v] = v^T (p^k)^T p^j v = \text{some huge mess and not just } \delta_{jk}$$

An analogous situation would be to imagine two unit vectors in E^2 that are not orthogonal. The direct E^2 scalar product would show non-orthogonality, but my "induced" scalar product would show orthog. We are just associating each "position" in the column vector with a basis vector. I don't know what the official term for doing this is, I have used the word induced, maybe associated is better.

Math for the people, by the people.

[Encyclopedia](#) | [Requests](#) | [Forums](#) | [Docs](#) | [Random](#)

Login

[create new user](#)name: pass: [forget your password?](#)

Main Menu

sections[Encyclopædia](#)[Papers](#)[Books](#)[Expositions](#)*meta*[Requests](#) (74)[Orphanage](#) (3)[Unclass'd](#) (1)[Unproven](#) (274)[Corrections](#) (72)*talkback*[Polls](#)[Forums](#)[Feedback](#)[Bug Reports](#)*downloads*[Snapshots](#)[PM Book](#)*information*[Docs](#)[Classification](#)[News](#)[Legalese](#)[History](#)[ChangeLog](#)[TODO List](#)

▲ proof of Jordan canonical form theorem

(Proof)

This theorem can be proved combining the cyclic decomposition theorem and the primary decomposition theorem. By hypothesis, the characteristic polynomial of T factorizes completely over F , and then so does the minimal polynomial of T (or its annihilator polynomial). This is because the minimal polynomial of T has exactly the same factors on $F[X]$ as the characteristic polynomial of T . Let's suppose then that the minimal polynomial of T factorizes as $m_T = (X - \lambda_1)^{\alpha_1} \dots (X - \lambda_r)^{\alpha_r}$. We know, by the primary decomposition theorem, that $V = \bigoplus_{i=1}^r \ker((T - \lambda_i I)^{\alpha_i})$. Let T_i be the restriction of T to $\ker((T - \lambda_i I)^{\alpha_i})$. We apply now the cyclic decomposition theorem to every linear operator $(T_i - \lambda_i I): \ker(T - \lambda_i I)^{\alpha_i} \rightarrow \ker(T - \lambda_i I)^{\alpha_i}$. We know then that

$\ker(T - \lambda_i I)^{\alpha_i}$ has a basis B_i of the form

$$B_i = B_{1,i} \cup B_{2,i} \cup \dots \cup B_{d_i,i} \text{ such that each } B_{s,i} \text{ is of the form}$$

$$B_{s,i} = \{v_{s,i}, (T - \lambda_i)v_{s,i}, (T - \lambda_i)^2 v_{s,i}, \dots, (T - \lambda_i)^{k_{s,i}} v_{s,i}\}.$$

Let's see that T in each of this "cyclic sub-basis" $B_{s,i}$ is a Jordan block:

Simply notice the following fact about this polynomials:

$$X(X - \lambda_i)^j = (X - \lambda_i)^{j+1} + X(X - \lambda_i)^j - (X - \lambda_i)^{j+1} =$$

$$(X - \lambda_i)^{j+1} + (X - X + \lambda_i)(X - \lambda_i)^j = (X - \lambda_i)^{j+1} + \lambda_i(X - \lambda_i)^j$$

$$\text{, and then } T(T - \lambda_i I)^j(v_{s,i}) = (T - \lambda_i)^{j+1}(v_{s,i}) + \lambda_i(T - \lambda_i I)^j(v_{s,i}).$$

So, if we also notice that $(T - \lambda_i I)^{k_{s,i}+1}(v_{s,i}) = 0$, we have that T in this sub-basis is the Jordan block

$$\begin{pmatrix} \lambda_i & 0 & 0 & \dots & 0 & 0 \\ 1 & \lambda_i & 0 & \dots & 0 & 0 \\ 0 & 1 & \lambda_i & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & \lambda_i \end{pmatrix}$$

? d_i ?

So, taking the basis $B = B_1 \cup B_2 \cup \dots \cup B_r$

, we have that T in this basis has a Jordan form.

This form is unique (except for the order of the blocks) due to the uniqueness of the cyclic decomposition.

end.

"proof of Jordan canonical form theorem" is owned by gumau.

(view preamble)

View style: ☐ HTML with images ☒  

This object's parent.

Cross-references: decomposition, blocks, order, polynomials, Jordan block, cyclic, basis, linear operator, restriction, factors, annihilator polynomial, minimal polynomial, characteristic polynomial, primary decomposition theorem, cyclic decomposition theorem, theorem

This is version 4 of proof of Jordan canonical form theorem, born on 2004-03-15, modified 2004-03-15.

Object id is 5709, canonical name is ProofOfJordanCanonicalFormTheorem.

Accessed 978 times total.

Classification:

AMS MSC: 15A18 (Linear and multilinear algebra; matrix theory :: Eigenvalues, singular values, and eigenvectors)

Pending Errata and Addenda

None.

Discussion

Style: ☐ Threaded ☒ Expand:  Order: ☐ Newest first ☒ 

No messages.

Interact

[rate](#) | [post](#) | [correct](#) | [update request](#) | [add example](#) | [add \(any\)](#)