Matthews MP274 Linear Algebra Notes (1991, web) PhL 12.20.04
finish 1.6.05

This is a very interesting set of notes that starts out fairly innocently, but then drills its way right into the heart of the Jordan Canonical form. The notes are rigorous and complete, everything is proven along the way, and the language is one I can understand. The notes are in a set of PDF files all of which I have downloaded. I have only printed out and read the first 4 chapters because these lead to the subject of the Jordan Canonical Form which is what I was interested in. These notes describe the Matthews dot diagram which I see mentioned nowhere else on the web, so I guess this nice graphic tool is his own invention?

Location of the notes: http://www.numbertheory.org/courses/MP274/

1. Linear Transformations.

I think many of these things were covered in my now-missing red Math 21 books. One main theme of this section is to keep in mind that a linear transformation (LT) is not the same as a matrix, but there are parallel universes going on here. These 19 pages let us understand the notation both in symbols and words that this author will be using in later sections.

Let's review this 19 page chapter now:

p1 defines an **LT** as **T:U→V**, it is a mapping in my book, and U and V are "vector spaces" the reader is assumed to know what a "vector space" means.

defines $\mathbf{V_n(F)}$ as an example of a possible vector space: n-dimensional column vectors
The **F** means the field F of the components of this column vector.
Keep in mind that F could be R, C or maybe a finite Galois field!

p 2 Define **Ker(T)** in the mapping world, whereas **N**(A) (nullspace) is in the matrix world.
Defines a **subspace**, see also Kazdan.
Theorem: Ker(T) is a subspace of U.
Define **Im(T)** as the **range** of T, that portion of V that gets hit by T(u) as u ranges over U.
Define $\mathbf{T_A}$ as the LT that we associate with the matrix A.
Define **C(A)** = the **column space** of a matrix A.

p 3 He should but fails to define **rank(T)** as dim(Im(T)).
He should but fails to define **nullity(T)** as dim(Ker(T)).
Theorem: **rank(T) + nullity(T) = dim(U).** He gives a detailed proof which I skipped for now.
Notice this is about the mapping T, not about matrices, and U is the domain of T.

p 4 **Theorem**: (**The Subspace Dimensionality Theorem**.)
$\dim(U \cap V) + \dim(U + V) = \dim(U) + \dim(V)$ where U,V are subspaces of W
I very dimly recall seeing this before, centuries ago.
Define: **direct sum** of two vector spaces, write as $U \oplus V$.

p 5 Theorem: $\dim(U \oplus V) = \dim(U) + \dim(V)$

p 6 Now we come to something very new to me. You really need to think of an n x m matrix as having matrix elements which are somehow related to the basis in U and the basis in V. You generate the matrix elements a_{jk} by looking at the action of T on each basis vector u_k in U. This gives a lincom of v_k as the result with the matrix coefficients. Therefore, the actual matrix elements *are specific* to your choice of the two sets of basis vectors!! If you change the basis in either U or V, your matrix changes. I am usually working with m = n and the same basis in both U and V. But in general, you should think this way:

$$A = [T]_{\beta}^{\gamma} \qquad x = [x]_{\beta}$$

where the single letter β represents a set of basis vectors in U, and γ in V. Similarly for the column vector x as shown, the elements of a column vector are basis-specific. So this is just a nice way to make this dependence very explicit. Notice that γ and β are *not* subscripts, they are part of the labeling of the matrix.

So this is a significant point being made. The mapping T is basis independent, whereas a matrix mapping T_A is basis dependent or specific.

p 7 Here we have a nice example where T(X) = [A,X] where A is a fixed 2x2 matrix. We find that the two basis vectors of ker(T) are I and A, so nullity = 2. We then compute a 4x4 matrix which is $[T]_{\beta}^{\beta}$ = B. My idea of using single-one 2x2 matrices is used as the basis, then flattened into a 4D col.

p 8 Define meaning of $\mathbf{T_1 + T_2}$ both mapping U → V

Define **Hom(U,V)** to be the vector space whose elements are all T that do T:U→V.

(I seem to recall this thing having a strange name, but it is lost).

Note: normally we think of U and V as the "vector spaces", but here we are thinking of T itself as a vector in a vector space. Stakgold never speaks in this manner. As for the origin of "Hom", the word homomorphism (same form) is sometimes used to mean a mapping between two sets where something is "same" in the second as in the first. In a linear map, perhaps f(a+b) = f(a) + f(b) shows that the notion of the operation "addition" has a parallel image in the second space. So just think of "homomorphism" = LT, so **Hom** is then just set of LT's which form a vector space.

p 9 How do you write Ax = y in our fancy notation here? See page 8 bottom $[T(x)]_{\beta} = [T]_{\beta}^{\gamma} [x]_{\beta}$. In this notation, remember that the β and γ are not subscripts. There is matrix multiplication implied on the right side, but the subscripts are hidden. The author is trying to show where "matrix multiplication" is coming from.

Composition of two transforms T_2T_1 and **matrix multiplication** for these guys.

p 10 The identify transform is written I_U since it maps U → U. We then get

$I_V T = T I_U = T$ // remember which space your I has to be in!

p 11 Now talking about S being T^{-1} . Notion of left and right inverse when spaces different.

p 12 An **isomorphism** is a mapping that is **1-1** (**injective**) and **onto** (**surjective**).

One to one: Obviously, T(u) = v maps you to only one point in V. The question is whether two points in U can map into the same point in V. If the mapping is one-to-one, the answer is no. There is a "one to one correspondence" between points in U and points in V.

Onto means that T(u) hits every element in V as u varies in U. In other words, Im(T) = V.

Lemma: T is 1-1 $\Leftrightarrow$ Ker(T) = {0}. { If not, let u1,u2 $\rightarrow$ v, then u1-u2$\rightarrow$0, so Ker(T) has u1-u2}

Theorem: in the m x n matrix world,

onto $\Leftrightarrow$ dimC(A) = m $\Leftrightarrow$ rows of A are lin indep

1-1 $\Leftrightarrow$ dimN(A) = 0 $\Leftrightarrow$ rank(A) = n $\Leftrightarrow$ cols of A are lin indep

Proofs are not given, I don't see offhand the rightmost parts of these claims, I am used to rows and columns being on the same footing since square matrices.

Theorem: if A is invertible, then the mapping is an isomorphism so you get 1-1 and onto.

p 13 **Theorem**: for an isomorphism T, you must have dim(U) = dim(V)

Theorem: Here we go up yet another level in terms of vector spaces. We now think of Hom(U,V) as a vector in some larger space, and in that space, Φ: Hom(U,V) $\rightarrow$ $M_{mxn}(F)$. So the mapping Φ is surely an isomorphism.

Theorem: If T is invertible then T is an isomorphism, and then dimU = dimV.

p 14 **Theorem**: if a matrix is invertible, must have m = n

Theorem: if dimU = dimV, then isomorphism $\Leftrightarrow$ onto $\Leftrightarrow$ 1-1, all three are the same.

p 17 Up to now, we have had our identity in the form $[I_V]_{\beta}^{\beta} = I_n$. When bases are the same, you get this matrix with 1's on the diagonal and you call it your matrix identity. However, suppose we make the two bases be different (even though both are in the space V). Then we can talk about:

$$P = [p_{ij}] = [I_V]_{\beta}^{\gamma} = \text{change of basis matrix}$$

Notice that [.....] seems to be a way to write a matrix, or perhaps to list elements of a set. We then end up thinking of this thing as x' = Rx, a basis change.

p 18 Then we can ask about a basis change acting on a T. We get this interesting result:

$$[T]_{\beta}^{\beta} = P^{-1} [T]_{\gamma}^{\gamma} P \qquad \Leftrightarrow \qquad B = P^{-1}A P$$

and this is our famous **similarity** transformation. It is just a change of basis.

2. Polynomials over a Field

I have been long waiting for this section to appear somewhere. When you see "minimal polynomial" in the Jordan discussion, you know you are getting involved with many of the "things" in my Galois Fields book. But there, I was specific to finite fields, and here Matthews is going to be more general and attuned to where we are going.

p 20 Polynomial n-tuple notation with lowest powers first, so x = (0,1,0...) up to n items.

How to multiply them. Multiplication is associative so p(rs)=(pr)s which follows from normal multiplication being associative.

p 21 def(fg) = deg(f) + deg(g) and similar basic rules

p 22 . He shows here that the **Lagrange Multiplier** $L_i(x)$ which he calls $p_i(x)$ forms a set of basis functions in the space $\mathbf{P_n[F]}$ meaning polys of degree n or less over field F. **F[x]** means all polys of x.

p23. The first corollary 2.1 here shows how you can write any f(x) using this basis, and the coefficients are $f(c_i)$ where these were the N+1 points you used in your Lagrange functions. I think of this as an interpolation of the data. In Corollary 2.2 we see that if f(c) = 0 for all n+1, then f ≡ 0, which shows that no poly of degree n can have n+1 roots. Just a spin-off fact. The fitting poly is unique.

p24. Notation **f | g means f divides evenly into g**. We than have the idea that f = qg + r where degree of r is less than degree of divisor g, and q and r are unique, these are all polys. This simple fact is called **Euclid's Division Theorem.**

p25. Now comes the infamous **Euclid's Division Algorithm**, notice the different name. You use the Theorem above many times to find the greatest common divisor of two polys f and g. I finally see how this works, but I had to fiddle for a while. All new to me. You end up with d = gcd(f,g) such that d | f and d | g and no smaller d can do both. A corollary fact is that you can write d = uf + vg where u and v are some polynomials you can figure out.

Now this all has a parallel in the world of integers as opposed to polynomials. You can find the greatest common divisor of two integers in the exact same way. Planetmath has a nice example all done out where you have gcd(756,595) = 7 and 37*756 - 47*595 = 7 which is that second part. The integer version is the real Euclid in his Elements, first real algorithms described in that book in ~350 BC. The little extra fact goes by the name **Bezout's Lemma**.

p26. An **irreducible** poly is one that cannot be factored into smaller polys (for a given field). We than have a little example to ponder: [I use this same definition on page 3.4 of my Galois book.]

Lemma: f(x) = (x-a)q(x) + f(a) is the famous "**remainder theorem**", we can say (x-a) | f ⇔ f(a)=0.

Now comes our example which is $f(x) = x^2 + x + 1$ and he picks $F = Z_2$ the binary world where • means XOR. He notes then that f(0) = 1 and f(1) = 1 so f(x) ≠ 0 for every point in U! Thus, there can be no factor of the form (x-a) which of course would have to be x or (x-1). This would also be irreducible in F = R, but not in C.

p27. **Theorem 2.2:** suppose f does not divide g *and* f is irreducible. It seems clear that the only thing that can divide f is f and 1, but we have said f goes not divide g, so the only thing left is gcd(f,g) = 1. And then the "little extra fact" says 1 = uf + vg.

Corollary 2.4. Let f = irreducible and f | gh. Then f|g or f|h. I think if could divide both as well, but it is never possible that it divides neither g nor h, it must divide one of them for sure.

Theorem 2.3 You can "factor" any poly into a unique product of irreducible polys. I believe it.

Theorem 2.4 says that, for a given degree n, you can always find an irreducible poly of that degree, *but* for some reason the field has to be a finite field to make this work. The proof is very long and probably quite interesting and that is why author has written it all out. He does not comment on q being finite.

p 32. The **Minimum Polynomial of a Square Matrix** is very clearly defined. You keep raising to powers A^r and eventually you will find that A^r = sum of lower terms. This has to happen because the dimension of the space is n^2 and you can't keep building forever. If we have gone a while, then we know that 1, A, A^2 etc are all lin indep. But there cannot be more than n^2 lin indep basis "vectors". So we know for sure that we hit r before $r = n^2$. But from C-H we also know that $r \leq n$ because we know there will be the char poly. The min poly is called $m_A(x)$ and we know that $m_A \mid ch_A$ (note the notation for char poly that he uses!). We make things monic.

p 33. **The Min Poly Algorithm.** Just compute powers A^r and the lowest power than is a lin com is the answer. You know you will reach it before $r = n^2$ as just noted.

Theorem 2.6. If you give me any poly of degree n, I can give you a matrix for which that poly is the min poly! It is shown on page 33. For the matrix shown, he shows that f(A)=0 and that it is the min poly. But it is also the char poly for this matrix. He forgets to say it, but this thing C(f) is called the **companion matrix**, and he will use it later.

p 35 Exercise 2.2 shows us the now famous **Jordan Block** but in lower triangular. We are supposed to imitate the previous proof to show that the min poly for this thing is $(x-\lambda)^n$ and again this happens to be the char poly as well.

This is not trivial. If you just take powers J, J^2 etc, it is not obvious when you get a result that is a lin comb. BUT, you do know that $m_J(x) = (x-a)^r$ where $1 \leq r \leq n$. So consider powers of (J - aI) which has the form of an empty matrix with a diagonal of 1's on the first off-diagonal. So (J - aI) = D1, to make a very compact notation for this matrix. This matrix is not a lin com of J and I, so we looked at $(J - aI)^2$ = D2. But D2 is not a lincom. Keep going and you find that $(J - aI)^n = 0$ and then you are done and you see that the degree of the min poly is all the way up to n. The appearance of D1, D2... fits with my Hessenberg product theorem, by the way, in terms of matrix shape at least.

So this is a very key fact about J_n . It's min poly is the largest it can be! $(x-\lambda)^n$ That largest is of course the char poly.

Direct sum of matrices is defined and properties given.

The **Least Common Multiple** lcm(f1,f2,f3...) is the smallest poly f into which they all divide (evenly).

p 36. We are getting very close to home now, Matthews is aiming his many weapons. What is the min poly of a direct sum matrix? Answer is

Theorem 2.7 Min poly of a direct sum of matrices = LCM (the min polys of the individual matrices)

This is proven using the simple direct sum properties. Again, the thing on the right here is the smallest poly into which all the individual polys divide. Another part of the theorem says that you multiply the char polys, but that seems pretty obvious from the overall block diagonal form, I think I could make a formal proof for this in various ways.

p 37. Suppose f is a product as shown here of some irreducible polys with little exponents with some extra factor constant c. And suppose g has the same form, with different exponents and d. What can we say about GCD and LCM ?? GCD(f,g) is a poly which has to divide into both f and g, so each separate factor has to divide into both. Now what divides into both p_1^a and p_1^b . The SMALLER one of course. So the GCD (p_1^a , p_1^b) = $p_1^{\min(a,b)}$. On the other hand, LCM(f,g) has to be divided by both f and g, so for each term, need something into which both p_1^a and p_1^bcan divide, and that is going to be the *larger* one, so we have LCM (p_1^a , p_1^b) = $p_1^{\max(a,b)}$

So all the results here look very good and promising.

Example: Matrix with just diagonal λ_i . Use direct sum to consider then just the 1x1 matrices. Each has a min poly $(x-\lambda)$. Then use the LCM rule and you see that you get one factor for each eigenmanifold.

p 38. On this page, author claims he is going to create a finite field with p^n elements where p is prime and where f(x) is some irreducible polynomial of degree n over the field Z_p. Remember that Z_p is itself only a field if p is prime. The elements of this larger field are going to be n x n matrices. Not just any matrices, but those you get by starting with a polynomial g(x) with coefficients in GF[p] = Z_p (the modulo finite field of order p) and then you create your field-member matrices as g(A), where A is the companion matrix to some irreducible f(x) of the type of g(x). So finding such an irreducible f(x) of degree n is required before you can build A = C(f), but then you have it. The basis matrices of the field are $1,A,A^2 ...A^{n-1}$. Field elements are thus $aA^0 + bA + cA^2$ etc, where the coefficients are in Z_p. There are only p such coefficients, so we can make $p*p*...*p = p^n$ such matrices, and these are then the p^m elements of this finite field we are constructing.

p 40 The number p is called the **characteristic**. We then learn that in general, if $q = p^n$ = size of your field, then for *any* x in the field, $x^{q-1} = 1$. We don't prove this, but "it can be shown". A special case when n = 1 and you have $x^{p-1} = 1$ for p = any prime, so this is then true for x = 1,2,3,4,5...p-1 . $x^p = x$ is another way to say this. This fascinating number fact is called **Fermat's Little Theorem**.

p 41 On this and the next page, author makes the usual parallel connection between T and A. So any LT T also has a char poly and a min poly. Good to show, but the results are forgone conclusions.

p 42 Remember that a **ring is a field without inverses**. We define two different rings, and then we show that they are in 1-1 relationship. The first ring consists of matrices *whose elements are polys* of degree n, an example is shown bottom of page 42. You can expand this out as a poly with matrix coefficients as shown. The second ring top page 43 is very similar, it is just the set of *polys with matrix coefficients*. The difference is a bit technical and is reflected in the field symbols. I don't think the distinction will be used in later parts of the author's notes.

p 43 We are not surprised to learn that in the world of polynomials with matrix coefficients, you have a **Euclid Remainder theorem for matrices**. That theorem you recall says f(y) = (y-a)q(y) + f(a). We now restate the same result, but think of the coefficients of y as matrices, so we are talking the second ring above. We can cap things to remind us they are matrices: F(y) = (yI-A)Q(y) + R(A). There is an issue here with which side you put certain matrices in R, and in this version powers of A are on the left. Only R has products of matrices, so that is where we have to worry about left and right.

p 44 Remember that an irreducible f poly is one that cannot be factored. This theorem makes this claim:

$$f \mid ch_A \Rightarrow f \mid m_A$$

This result is not really obvious, but it seems very reasonable to me. Two proofs are given from different sources, but I have not studied either one.

p 45 Suddenly we have a change of topic. He defines the word **diagonable** to mean diagonalizable by a similarity. Then come some important theorems.

Theorem 2.12: A diagonable => $m_A = (x-c_1)(x-c_2)...$ // all to first power, where c_i are the distinct diag el's.

My proof: m_A is invariant under similarity, so look at the diagonal form. Think of it as a direct sum of blocks for each distinct c_i. For each block, we know that char poly = $(x-c_i)^a$ where a = alg mult, but we know also that min poly = $(x-c_i)$ because the block B is just a multiple of I_a and then $B^1 = c_i I$ so we know that the min poly is degree 1. Now combine all the blocks and use the direct sum theorem we had which says find LCM (of all the min polys) and that is the result shown above.

Corollary: Since J_n does *not* have $m_A = (x-c_1)(x-c_2)...$ which would be $(x - c_1)$, it cannot be diagonalized!

p 46 **Theorem 2.14**: Geo Mult ≤ Alg Mult, or in Stakgold terms or mine, $m \le k$. A nice proof is given.

p 47. **Theorem 2.15**: Now comes a pretty fancy theorem. It really says this: if each eigenmanifold has no defect, meaning each one has geo mult = alge mult, then the nullspaces of $(T - c_iI)$ partition the full space of your matrix V, they are of course independent, and there is nothing left over. In this case, we can write the full space as a direct sum of the nullspaces. In each nullspace, since we have the full complement of eigenvectors, we know we can diagonalize using X formed from those eigenvectors, so let's do so. Then the entire matrix becomes diagonal. This is what the final line says on page 48.

p 49. Here we have an example with two nullspaces which have full rank and sure enough, the matrix is diagonable.

p 50. **Theorem 2.16** Another very new to me theorem. It generalizes Theorem 2.12 to both directions, so now we have: [diagonable $\Leftrightarrow m_T = (x-c_1)(x-c_2)...$] We are talking matrix T and distinct eigenvalues c_i in this discussion. It then goes on to show that, in this case, if you define $T_i = L_i(T)$ where L_i is the Lagrange Multiplier encountered earlier (and called p_i), then these T_i have interesting properties.

$\Sigma T_i = I$ T_i are the **principle idempotents** of T

$\Sigma c_i T_i = T$

$T_i T_j = T_i \delta_{i,j}$

These things look like projection operators to me. I went through the proof of everything here on p 51.

p 52 The corollary here just says that if your char poly has all first powers, then you are of the right form to be diagonable. The example that follows shows that for 2x2, both the symmetric and skew-symmetric matrices are diagonable.

Comments: I think the tools needed for deriving the Jordan Normal Form are mostly assembled by this point. At least the subtools needed to make the final tools. In the next chapter, we are going to see one of the Big Tools known as the Primary Decomposition Theorem.

3. Invariant subspaces

p 53. We know what it means for subspaces V_1 .... V_t of V to be **independent**, this is restated here. It is the same idea as linear independence of basis vectors. If you can write $\Sigma_{all} v_i = 0$, then if the v_i are *independent* column vectors, say, then all $v_i = 0$. If not, then you have at least one as lincom of others, and that is not possible. So in the case of subspaces, you write $v = \Sigma_{all} v_i$ with v_i in V_i and since there cannot be a relation between them, all $v_i = 0$ if $v = 0$.

In general, when the spaces are independent and span the parent space, you have a direct sum and you can union the bases to get a basis for V. And of course the dimensions all add up.

The **T-invariant subspace** is quite obvious.

p 54. We have already seen (p 32) the min poly *of a matrix*. We saw that you can express any power A^r in terms of A^0, A^1 up to A^{s-1} where s is the degree of that min poly. We did not talk much about these powers A^n spanning any kind of space, but we can see they are independent. When we talked min poly for matrices, we meant that $m_A(A) = 0$ as a matrix, so certainly $m_A(A)v = 0$ when you apply this to any vector v in your space.

Now we define a less restrictive polynomial. We write $m_{A,v}(A)v = 0$. This poly $m_{A,v}(A)$ is the smallest poly that annihilates *this specific vector v*. It need not annihilate other vectors in V the way that we know $m_A(A)$ does. This is the **min poly of a vector v**.

What can be said about such an animal? We know of course that $m_A(A)v = 0$ for all v, but for our specific v we know that $m_{A,v}(A)v = 0$. If you take ANY f(A)v with f(A)v = 0, we know that $m_{A,v}(A) \mid f(A)$ because our guy is supposed to be the smallest one. In particular, then, $m_{A,v}(A) \mid m_A(A)$. So this new thing is a factor of m_A if not the whole thing.

Next concept **Theorem 3.2**: pick some v in V and any old f(x) you want, this forms a set of operators f(T). The claim is that the set of vectors of the f(T)v form a T-invariant subspace inside V. The subspace is spanned by the following set of vectors: v, Tv, T^2v ..... $T^{k-1}v$ where k is the degree of the min poly of vector v, so the dimension of this subspace is k. It looks so similar to the matrix concept above, where we had matrices A^0, A^1 ..... A^{s-1} where s was the degree of the min poly of A. But now we have vectors, not matrices. You "get to" the next vector to the right by applying another T, so we have the idea of **cyclic**. That is, basis = $(1,T,T^2,T^3...)v$. We cycle around the basis by applying T. So this thing is called the **T-cyclic Subspace** and is written $\mathbf{C}_{T,v}$ where C reminds us of Cyclic. Finally, the basis vectors are v, Tv, T^2v ..... $T^{k-1}v$.

At this point I find the bra-ket notation to be useful. We know that $\langle v|Tv\rangle = 0$ because I think the basis vectors are in fact orthogonal. He does not mention this, but it must be true. If I assume this is true, then we can create the following square matrix: $C_{ik} = \langle T^i v \mid T \mid T^k v \rangle$ which is just the matrix elements of the operator T with the same basis on both sides. It is then easy to show that this matrix is exactly the companion matrix we found earlier, but here the relevant poly is the min poly for vector v. This is shown on the bottom of page 56.

p 57 **Theorem 3.3.** Now apply the previous theorem but apply if to T' = T - cI. In this case, we expect the powers of T' to be our basis (they are shown top of page 57), and we will again assume orthogonal in

the same sense as before. The min poly for T is now assumed to be $M_{T,v}\ v = (T - cI)^k\ v$. Then the min poly for T' is $M_{T',v}v = (T')^k$. When we use this poly to make the coefficients in the companion matrix, we get $a_k = 1$ but all others are zero. The companion matrix for T' will then be as shown on page 56 with the right column all zero. But that then means that the matrix for T = T' + cI is exactly $J_k(c)$ as claimed. I wish he would have said some of these words I am now having to say!

I am going to skip the generalization which gives the hypercompanion matrix.

The **C-H Theorem**. I read through this and it looks good to me.

p 58. **Theorem 3.4**: **The Nullspace Nesting Theorem.** This is a very messy theorem with a long 2 page proof. To simplify the proof, we invent the more compact notation shown at the bottom of page 58, which is fine by me. I have been doing that a little already since I think matrix.

Now what exactly is this theorem trying to tell us? You assume that the min poly for the matrix T (we are done I guess with min poly for vectors) has the form of a product of monic polys raised to powers, where each of these polys is a (distinct) irreducible over our field F, whatever it is. The polys are called p_i and the powers b_i. I can live with that.

Now the first (implied) claim is that the *char* poly for our fancy matrix T whose *min* poly is this messy thing is in fact of identical form, but instead of exponents b_i, there will be some other exponents a_i. However, this theorem is not going to tell us what these a_i are, we have to wait for some later theorem to learn that.

OK, we can live with that deferral. We still have not stated the claims of this theorem. They are:

Claim #1: You can consider the nullspace of the operator $p_i^{bi}(T)$ which is one of the factors in our min poly. You can ask: what is the nullity of this nullspace? The answer is that $\upsilon(p_i^{bi}(T)) = a_i$ * degree of p_i .

Claim #2: We can consider a whole set of nullspaces where we replace b_i with lower integers ranging from b_i down to 1. The claim is that these nullspaces are nested in the Venn diagram sense, where the largest outer one is $\text{Ker}[p_i^{bi}(T)]$ and the innermost one is $\text{Ker}[p_i(T)]$. The nullities of these spaces can then be written as a long inequality as shown, namely, $\upsilon(p_i(T)) < \upsilon(p_i^2(T)) <$ etc. We have in Claim #1 an expression for the nullity of the largest nullspace. Another part of the claim which is not clearly stated is this: as you consider powers larger than b_i , the nullspaces stop increasing and stabilize in size. For example, $\text{Ker}[p_i^{bi+1}(T)] = \text{Ker}[p_i^{bi}(T)]$ and the nullities are the same. I think this part of the claim will be used later somehow to "compute" the value of b_i .

Claim #3: We know of course that rank + nullity = dim(V). Thus, in the sequence of nullspaces of the previous claim, as the nullspaces get larger, the ranges get smaller (rank = dim range, AKA the image). This results in a Venn nested set of ranges as shown as item (a). He does note that we don't therefore have to prove this claim, so he will only prove Claim #2.

Proof of Claim #2

(i) In this section, he shows that K p contains at least one element, and that therefore $K\ p \supset \phi$

(ii) We always know that $K(A) \subseteq K(BA)$ by drawing a trivial Venn diagram of these two nullspaces. This fact is used many times in Matthew's proofs. It is the "other direction" that is usually the problem of a proof. So here we trivially have that $K\ p^b \subseteq K\ p^{b+1}$ (a "containment") and he then proves that $\supseteq$ is also

true so that we end up with equality. In this proof, as in many other proofs, we use that famous Bezout Lemma that goes with the GCD theorem (Euclid's Division *Algorithm*). This little gem is of high value. So at this point, we know that $K\ p^b = K\ p^{b+1}$.

(iii) Now he uses an induction proof to continue this equality to the right, for example, to $K\ p^{b+1} = K\ p^{b+2}$. He wants to prove the item I have labeled (a), but he really only has to prove (b) because "the other direction" of the two containments is trivial. So he proves the second item (b), and thereby proves (a) and then we have things done "on the right" as far as you can see.

(iv) The only missing part now of Claim #2 is the interior section with, for example, $K\ p^{b-1} \subset K\ p^b$ where now b is any integer from 0 to b_i . This is "the trivial direction" (smaller contained in larger), but we have to rule out the = part of $\subseteq$. To do this, he finds an element in $K\ p^b$ which is NOT in $K\ p^{b-1}$.

Comments: So at this point we have proved his complicated item (b) on page 58, part of my claim 2. Now when you say something like $K\ p^{b-1} \subset K\ p^b$ without the equality part, the nullspace on the left is definitely "smaller" than the one on the right, and this means that υ(left) < υ(right). If the nullities are equal, the spaces are the same size. So we have now proven this fact which will find use later:

$$0 < \upsilon(p) < \upsilon(p^2) < ..... < \upsilon(p^b) = \upsilon(p^{b+1}) = \upsilon(p^{b+2}) = ...$$

where b is the exponent of p in the min poly of matrix T.

Proof of Claim #1? This claim says that $\upsilon(p^b)$ = a deg p, but we are NOT proving that yet in this theorem, it is coming in a later theorem. However, if we accept this as true, then we can make the following observation: Suppose we start computing $\upsilon(p^h)$ for various powers h = 0,1,2... . I don't yet know exactly how we would compute these things, but suppose we could. Then when h reaches b, we get $\upsilon(p^h)$ = a deg p. We of course know that a is the alge mult and we will have deg p = 1, so we will know when we hit this value. We can then deduce that the value of h where we first get a deg p must be h = b, and that is a way in which we can "compute" b. Remember that b_i is the exponent of $p_i(\lambda) = (\lambda - \lambda_i)$ in the min poly for matrix T.

Comments on finding the min poly of a matrix T.

The above theorem gives us a method to find this min poly, so finally I understand his title for the section in which the theorem appears on page 58. We know that the min poly has to divide the char poly, so we know the general form of the min poly, we just need to find the exponents b_i. One way is to "search" by trying all possibilities, sometimes this is reasonable, you start with $b_i = 1$ and then raise it, doing them all of course at the same time in your search. Another way is to keep raising matrix T to higher powers and check each time whether you have a lin comb of lower powers. This does not sound very easy if you have a 10 x 10 matrix -- how would you know without lots of work that you don't have a lin comb of previously computed matrices?? So I think the method given in this section will have practical use.

p 61 **Theorem 3.5:** (**The Primary Decomposition Theorem**)

We start as with the previous theorem with the min poly of T being a product of p_i^{bi} . Here are the claims.

Claim A: The nullspaces Ker[p_i^{bi}(T)] for each i (for i = 1 to t) partition the whole space V in which T acts, so you can write the whole space B as a *direct sum* of these nullspaces.

Claim B: Each nullspace Ker[p_i^{bi}(T)] is a T-invariant subspace.

Claim C: $\upsilon(p^b)$ = a deg p. This item was mentioned in the theorem page 58 and is here proven.

Proof: We do things in this order: Claim A, Claim C and then Claim B.

(i) First, we define operators T_i as $(f_iq_i)(T)$ where q_i is "the rest" of the min poly when you remove p^{bi}, and where f_i is any poly. Then using our old friend Bezout, we show that these T_i add up to the identity matrix, and moreover, that they satisfy $T_i\ T_j = T_i\ \delta_{i,j}$ and $T_i^2 = T_i$. So these "idempotent" operators certainly look like useful "projection" operators that might define some subspaces!

(ii) Now we consider the *ranges* of the operators T_i . We want to show that $V = ImT_1 + ImT_2 + ...$ is in fact a partition of the space V, so we can write as a direct sum. We want to show that these spaces are independent subspaces in the sense of the discussion on page 53. As shown on that page, we need to show first that every element v in V has a unique "partition" into v_1 in ImT_1, and v_2 in ImT_2, and so on. And we have to show the independence idea.

So how does Matthews do this? First, he shows that we can write $v = Iv = (T_1v + T_2v + ...) = v_1 + v_2 ...$, so we have the decomposition of $v \in V$ into its pieces in the subspaces. He then shows the independence idea by assuming that $v = 0$, and then showing that this implies that each $v_i = 0$. Matthews is a little unclear in his proofs with his use of the word "For", and in his general logic of presentation. For example, on the bottom of page 61 he states the direct sum result, but it is unclear whether he thinks he has proven it or is about to prove it. The "For" at the top of page 62 is completely out of place. This is just a little fact that gets used in what follows. But OK, I agree with it all and I agree that we have proven that V = direct sum of ImT_i , as shown bottom of page 61.

(iii) Here he shows that $ImT_i = K\ p^b(T)$. He does this very cleanly by showing $\subseteq$ and then $\supseteq$ so =. Nothing fancy is going on here, it is all very clean. So we have at this point **proved Claim A** in full!

(iv) Now we have some cloudiness. We identify $V_i = ImT_i = K\ p_i^{bi}(T)$ which is fine. Now comes the tricky part. Think of writing the operator T as a direct sum of operators, so we have this parallel:

$$V = V_1 \oplus V_2 \oplus V_3 ...$$

$$T = L_1 \oplus L_2 \oplus L_3 ...$$

The first line is a direct sum of spaces, while the second line is a direct sum of operators. Matthews has in fact used this notation on page 35 where he talks about a direct sum of matrices, and he gives properties there. Then on page 36 he states the fact that to me is pretty obvious by now that the char poly is the product of the char polys of the submatrices in the direct sum, Theorem 2.7 page 36.

So I have two gripes at this point. (1) he has already used the notation T_i for the idempotent operators, so he is forced to use some new letter L_i for the submatrices of T. (2) He fails to refer to his earlier work for example when claiming the product for the char poly on page 62 bottom.

Now we know that $p_i^{bi}(T)v = 0$ for ALL v in V_i, so identically we can say $p_i^{bi}(L_i) = 0$ in this subspace. In this subspace V_i for matrix L_i we have both a char poly and a min poly. Since we have just shown that $p_i^{bi}(L_i) = 0$, it seems reasonable to assume that both the char and min poly have this form with some unknown as yet exponents. So let e_i be the min poly expo, and let d_i be the char poly expo, both going with this submatrix L_i. We know by assumption that ch_T = powers raised to a_i and m_T = powers raised to b_i. So

(A) ch_T = powers raised to a_i = product of powers raised to d_i , so at once $a_i = d_i$. In other words, for the submatrix L_i we have char poly = p_i^{ai} . This all seems reasonable because L_i is associated with a particular eigenvalue that is part of the $p_i = (x - \lambda_i)$, and a_i is the alge mult of this eigenvalue, and this is what you always find in a char poly -- the expo is the alge mult.

(B) m_T = powers raised to b_i = product of powers raised to e^i , so at once $b_i = e_i$. In other words, for the submatrix L_i we have that min poly = p_i^{bi} . We are not surprised by this result!

(C) What is dim V_i ? it is the degree of the char poly of L_i which is $(\lambda - \lambda_i)^{ai}$ in our case, but more generally we might have more than a linear poly for p, so it is degree p^a which is in fact a deg p, and we have **now proved Claim C.**

What about **Claim B** that Ker[$p_i^{bi}(T)$] is a T-invariant subspace? Here is my own proof of this. Suppose $w \in$ Ker[$p_i^{bi}(T)$]. Then we know that $p_i^{bi}(T)w = 0$. Now consider w' = Tw. Clearly we can say

$$p_i^{bi}(T)w' = p_i^{bi}(T)(Tw) = (p_i^{bi}(T)T)(w) = (Tp_i^{bi}(T))w = T\,(p_i^{bi}(T)w) = T\,(0) = 0$$

Therefore, if $w \in$ Ker[$p_i^{bi}(T)$], then $Tw \in$ Ker[$p_i^{bi}(T)$] as well, so Ker[$p_i^{bi}(T)$] is a T-invariant subspace.

So what does this theorem really say? If the min poly of a matrix T has the specialized form shown, then the nullspaces of the factors of that min poly partition the whole space and you have a direct sum.

Now let's guess the future. We know that the char poly of a matrix T can be factored into factors $(x-a)^k$ at least if the field is F = C. This is usually thought of as $(\lambda-\lambda_1)^k$ where k is the alge mult [the a_i]. The min poly we suspect has a similar form with perhaps lesser exponents [the b_i]. Thus, we can think of the matrix T as a direct sum of a nullspace for each factor, that is, for each distinct eigenvalue. This is going to mean that there is some basis choice in which we can get our matrix T into that Jordan Normal form. Within each block, we will perhaps take the basis shown on page 57. Then I think we will find that certain 1's are missing if you don't have full geo mult, and each of those blocks can itself be reduced to a sum of smaller blocks.

p 63. Theorem 3.6 says that if two matrices **commute**, you can diagonalize them both at once in some basis. OK. Then he states **Fitting's Lemma**, but I do not see the significance of it and no proof is given, I have the impression he just stuck this in here to have it somewhere.

So now we are finally ready for the next section:

4. The Jordan Canonical Form

We need **lots of prep work for this section** that Matthews has not provided, but I have found it elsewhere, luckily! www.math.mcgill.ca/labute/courses/270/add.pdf

Comments on these other notes:

Theorem 1 says this: $\dim(U) = \dim(\text{Ker}(T)) + \dim(\text{Im}(T))$ for $T: U \to V$

This is just the famous result that N = nullity + rank, expressed in different notation. Remember that rank is the dimension of the range of T, and $R_T = \text{Im}(T)$ and $\text{Ker}(T) = N_T$ the nullspace.

Corollary 2 is an interesting application of this theorem, but to understand it, you have to draw a fairly complex picture which I have done in pencil, call it Figure 1. We are going to talk about a composite mapping. First we have $T: U \to V$ and then $S: V \to W$, so the composite is $ST: U \to W$. There are many things needing words here:

(1) I have drawn Ker(S) in the V space, the set of points which map to 0 in W. This contains the 0 of V.
(2) Next, also in the V space I have drawn Im(T).
(3) Next, in the U space I draw Ker(ST) and, within it, Ker(T). Remember that the "smaller" kernel is the one with less operators, so $\text{Ker}(T) \subseteq \text{Ker}(ST)$.

Now we have to draw various mapping arrows.

(4) Points within Ker(ST) in U have to end up at 0 in W on the far right, but on their way, just applying T, we have to land in the intersection region of Im(T) and Ker(S). The reason is this: any action of T on U (call it Tu) must take us to a point within Im(T) by definition of Im(T). But in order to end up in 0 in W, we have to have Tu lie also in Ker(S) so that S(Tu) = 0. So this is how we arrive at the intersection region written as $\text{Im}(T) \cap \text{Ker}(S)$. [That was a big question I had, now I see where it is coming from.]

(5) Meanwhile, we can draw an arrow showing how Ker(T) in U maps into the 0 in V, and this 0 lies within the above mentioned intersection region.

(6) Now we "chop away" the rest of space U and define U' = Ker(ST) and we define $T_0: U' \to V$. This is just a restriction of T to a smaller domain. We now make these two claims:

a) $\text{Im}T_0 = \text{Im}(T) \cap \text{Ker}(S)$ [because you can no longer land outside this region in V]
b) $\text{Ker}T_0 = \text{Ker}T$ [because the entire KerT lies inside U', so T and T_0 are the same in this regard]

(7) Now we finally come to the result we want: First, our original rank + nullity law says this:

$\dim(U) = \dim(\text{Ker}(T)) + \dim(\text{Im}(T))$ for $T: U \to V$

Now apply this rule not to U, but to U' and we have

$\dim(U') = \dim(\text{Ker}(T_0)) + \dim(\text{Im}(T_0))$

Therefore we have proven this fact: (I make up a theorem number and name)

Theorem 2: (**ST Intersection Theorem**) For T: U→V and S:V→W, we know that:

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

which we can rewrite as

$$\dim(\mathrm{Im}(T) \cap \mathrm{Ker}(S)) = \upsilon(ST) - \upsilon(T)$$

Now let's apply this theorem to S = p(T) and T = p^{h-1}(T) and we get:

$$\upsilon_{h,p} \equiv \dim N_{h,p} = \dim(\mathrm{Im}(p^{h-1}) \cap \mathrm{Ker}(p)) = \upsilon(p^h) - \upsilon(p^{h-1})$$

and this concludes the section called Definition 4.1!

Notice that $\upsilon_{h,p}$ is not really a nullity, it is just a symbol. And we have given this name to the intersection space whose dimension is,

$$N_{h,p} = \mathrm{Im}(p^{h-1}) \cap \mathrm{Ker}(p)$$

Question: What is this space $N_{h,p}$? Matthews says nothing to this question. However, we can apply our earlier ST picture, and think about the total ST = p^h and we break this mapping into S = p^{h-1} and T = p, so T just performs that "last step" which actually takes us to the "0" point in V. Overall we are interested in Ker(p^h) and we are certainly allowed to think of this as Ker(ST) with S,T as just given. [Don't confuse the T of ST with the matrix T we are working with, which appears in p(T) for example, and which we are always not writing. These are completely different T symbols.] In the next theorem, we are going to show that this intersection space $\mathbf{N_{h,p}}$ gets smaller and smaller as h increases from h = 1 to h = b, and for h > b, the space shrinks to nothing.

I have drawn this situation in pencil Figure 2, keep with this document please!

The Basis Business. This is another key item in this section that I neglected on first reading! He says this: suppose you have a basis for Ker(p^h) which is $\langle u_1, u_2 ....\rangle$. We can map each of these with our first mapping in Figure 2 to get a set of vectors $\langle p^{h-1}u_1, p^{h-1}u_2 ...\rangle$ in our hatched intersection region in V. As noted on top of page 3, it is not automatic that $\langle p^{h-1}u_1, p^{h-1}u_2 ...\rangle$ is a basis in our hatched space. [I have just realized that he uses the notation A = <a,b,c> to say that "the space A is **spanned** by vectors a,b,v.] So at least we know that the set of vectors $\langle p^{h-1}u_1, p^{h-1}u_2 ...\rangle$ spans our hatched region $N_{h,p}$.

Now later he is going to use this idea and he will claim that this set is in fact a basis for $N_{h,p}$.

Theorem 4.1 (Nested $N_{h,p}$ Theorem) [p 65] The spaces $N_{h,p}$ have this nested containment property:

$$N_{1,p} \supseteq N_{2,p} \supseteq N_{3,p} \supseteq \ldots\ldots \supseteq N_{b,p} \neq \{0\} \quad \text{and } \{0\} = N_{b+1,p} = N_{b+2,p} = \ldots.$$

$$\upsilon_{1,p} \geq \upsilon_{2,p} \geq \upsilon_{3,p} \geq \ldots\ldots \geq \upsilon_{b,p} \geq 1 \qquad \text{and } 0 = \upsilon_{b+1,p} = \upsilon_{b+2,p} = \ldots$$

where the second line is about the dimensions of the spaces.

Proof: Well, the second line above follows from the first, so we only have to prove the first. We go back to page 58 (a) and recall that nested image theorem stuff which allows us to write Im $p^{h-2} \supset$ Im p^{h-1}, for example (part of the chain of such things). When we intersect this with Ker p, we conclude that

$$\text{Im } p^{h-2} \supset \text{Im } p^{h-1} \Rightarrow \quad [\text{Im } p^{h-2} \cap \text{Ker } p] \supseteq [\text{Im } p^{h-1} \cap \text{Ker } p] \Rightarrow \quad N_{h-1,p} \supseteq N_{h,p}$$

where we pick up "equality" for the case that there is no intersection at all (something like that).

Theorem 4.2 (The $N_{h,p}$ nullity sum rule.)

$$\upsilon_{1,p} + \upsilon_{2,p} + \upsilon_{3,p} + \ldots\ldots + \upsilon_{b,p} = a = \upsilon(p^b)$$

Proof: We already know from above that

$$\upsilon_{h,p} = \upsilon(p^h) - \upsilon(p^{h-1}) \qquad \text{with special case } \upsilon_{1,p} = \upsilon(p) - \upsilon(p^0) = \upsilon(p) - \upsilon(I) = \upsilon(p)$$

since the identity operator has no nullspace (one way to think of it). As he notes, you then get a simple telescopic cancellation which proves the above rule. We know that $\upsilon(p^b) = a$ deg p from page 63 top, but we are assuming that deg p = 1 now.

Example: Here we imagine that b = 4 and we are given the nullities shown for the powers of p, and from these we compute the nullities $\upsilon_{h,p}$ as shown. As promised, they form a decreasing sequence (theorem 4.1 second line). In this example, since we know b = 4, we also know that $\upsilon(p^4) = 10$ is our value of a.

Matthew's Dot Diagram.

Is this "his" invention? Probably, I will look it up later. Each dotted box represents one "unit" of $\upsilon_{h,p}$ nullity. From the sum rule, we know that there will be a total of "a" boxes, in this case 10. We start at the bottom row and draw $\upsilon_{1,p}$ boxes, which is 3 in the sample. Then above that we draw a row with $\upsilon_{2,p}$ boxes, and so on until we are done. We know there only be "b" rows in the picture, with a total of "a" dots in all rows. We also know that the width of the rows decreases as you go up.

You should mentally associate each row of the dot diagram with one of the $N_{h,p}$ spaces. The bottom row has h = 1, and this is associated with the space $N_{1,p}$ = Ker(p) and this is the space we are very interested in since it's basis is the eigenvectors of the matrix T within the particular eigenmanifold we are now working in. Remember that the entire dot diagram only applies to a particular p_i^{bi} factor and we are ignoring momentarily all the other factors. That is to say, the dot diagram and all this discussion only applies to ONE of the subspaces of the primary decomposition. So we are really talking here about the secondary decomposition level. The next row up in the dot diagram is associated with $N_{2,p}$ and so on. The topmost row is for $N_{b,p}$ which is the largest non-empty nested hatched space.

Now we can think about the number of columns and the height of each column. In this example there are 3 columns with heights $e_1 = 4$, $e_2 = 4$ and $e_3 = 2$. Clearly, the *number* of columns is determined by the number $\upsilon_{1,p}$ but he just refers to this number of columns as γ.

Somehow not yet revealed, we are going to associate this dot picture with the Jordan block structure for this eigenmanifold. But so far, we have simply described the picture and how you make it, not what its significance is.

Theorem 4.3. The claim is that you can find vectors v_i in V such that the following set of vectors,

$$\{ p^{e1-1}v_1,\ p^{e2-1}v_2 , p^{e3-1}v_3 \ldots\ldots ,\ p^{e\gamma-1}v_\gamma \ \} ,$$

forms a basis for Ker(p) which, by the way, is our eigenmanifold of interest. Note that Ker(p) is the end of the line of the spaces $N_{h,p}$ and is in fact $N_{1,p}$ which has dimension $\upsilon_{1,p} = \gamma$. So at least I agree that there should be γ basis vectors in a basis, and that is what we show here. I forgot to comment earlier that this number γ is the geometric multiplicity of our eigenmanifold! The reason is that Ker(p) = N(T - λI) and recall that our eigenvalue problem is Tv = λv or (T - λI) v = 0, so the number of eigenvectors is equal to the size of N(T - λI) which is Ker(p) and this size is γ.

Proof: His proof is pretty useless, he is only providing a dim hint. The number of boxes in a row is the number of basis vectors one would need *for the corresponding N space*, that is true. So in the example, the top row has $\upsilon = 2$ boxes, so we need 2 basis vectors. He just claims without proof what the theorem is stating, so this is not a proof. How do we know these sets of vectors form bases for the various spaces? I am going to have to look elsewhere! I wonder if he just ran out of time?

Second attempt at a proof. I have now looked at some examples below, and learned the span notation, and wrote the "Basis business" thing above, so maybe I can do it now. The top row in his example is associated with the hatched space $N_{4,p}$ which we know has $\upsilon_{4,p} = 2$ (two boxes in the top row). This is the dimensionality of the space $N_{4,p}$ (whether or not you refer to $\upsilon_{4,p}$ as a "nullity"). So we know that there are going to be two basis vectors for this space. We are now looking at our Figure 2 with h = 4. On the left we have Ker(p^4) which has dimension $\upsilon(p^4) = 10$ in our example on page 66. The basis of Ker(p^4) we could regard as being $< u_1,u_2,u_3.....u_{10}>$ as on page 65 top. We could then make a spanning set for the cross hatch region $N_{h,p} = N_{4,p} = < p^3u_1,p^3u_2,p^3u_3.....p^3u_{10}>$. However, we know that $\upsilon_{4,p} = 2$, so only two of these spanning vectors can be linearly independent, not all 10. Let's call the "right ones" v_1 and v_2 (perhaps $v_1 = u_9$). Then our basis for the space $N_{4,p}$ is $< p^3v_1,p^3v_2 >$, and this is what he has stated next to the top two boxes.

Now the space $N_{3,p}$ also has dimension = 2 in our example, so "things have not gotten larger". It must be that $N_{4,p} = N_{3,p}$, so this same basis $< p^3v_1,p^3v_2 >$ I guess works for $N_{3,p}$. But $N_{2,p}$ has dim = 3, so our hatched space has now grown. Because the spaces are nested, I think we can use the two vectors we already found, $< p^3v_1,p^3v_2 >$, and now we add a new one of the form pv_3 . Here we have power p^1 because that is the value of h-1 = 2-1 = 1 that we use in our "basic business" paragraph discussion above.

So it would appear that you start at the top, make a basis using some powers, then you move down to the next place it gets wider and you add one or more lower power "extension" basis vectors at that point, until you get all the way to the bottom, at which point you will have a set of basis vectors for the complete eigenmanifold.

Now look at the claim of this theorem in relation to the example. We used 3 in $< p^3v_1,p^3v_2 >$ because we were working first with $N_{4,p}$ having h = 4. The number of basis vectors with "3" is the number of columns of height 4. But in our example, we just have $e_1 = 4$, and $e_2 = 4$ so these account for the first two basis vectors having 3. Then we get to $e_3 = 2$ and that gives us the pv_3.

So it is slowly dawning on me. Once you have the dot diagram, this theorem is telling you HOW to make a basis for your eigenmanifold Ker(p).

Status: I am now "content" with Theorem 4.3 [at first blush, it was completely meaningless] . So now finally I can dive into the grand finale, which is Theorem 4.4.

Theorem 4.4. (**Secondary Decomposition**). This is the second stage where we shall write out an eigenmanifold as a direct sum of subspaces all having the same eigenvalue, and these will be the Jordan block set for given λ. Everything claimed here is specific to a particular eigenmanifold of T, which eigenmanifolds you can identify by looking at char poly of T.

Claim A: To each column of the dot diagram of height e_i we can associate a vector v_i (in V) and we can associate a min poly generated by v_i which is simply p^{ei} . We write $m_{T,vi} = p^{ei}$.

Claim B: The space $\text{Ker}(p^{bi})$ can be written as a direct sum of cyclic subspaces $C_{T,vi}$ which have dimension e_i. These are of the form discussed on page 57: $m_{T,v} = p^k$ (where $k = e_i$), Jordan basis = $(v,pv,p^2v...p^{k-1}v)$, and in this basis the submatrix $C_{T,vi}$ is $J_{ei}(\lambda)$. The *number* of cyclic subspaces equals the number of columns in the dot diagram, and the *dimension* of each is the height of its column which is e_i. This decomposition of $\text{Ker}(p^{bi})$ is called the "secondary decomposition". The "primary decomposition" was writing the total space V as a direct sum of subspaces of the form $\text{Ker}(p^{bi})$.

Claim C: (added by me). Once you have the secondary decomposition direct sum given above, there is an algorithm such that for each column of the dot diagram you can find an associated eigenvector of the form $p^{ei-1}v_i$. That is, this algorithm tells you how to find v_i for each column. The number γ of such eigenvectors equals the number of columns, which of course is the number of boxes in the bottom row, which in turn is the dimension of the nullspace Ker(p), so these are then the "true eigenvectors" of our eigenmanifold. This number γ is thus the geometric multiplicity of the eigenmanifold. The γ eigenvectors form a basis for the space we associate with the bottom row which is $N_{1,p} = \text{Ker}(p)$. The set of vectors v_i contains one vector for each column, and this vector is the generator of a T cyclic subspace for that column. We can compute the other subspace basis vectors using the powers rule. This we end up with this:

v_1	$\beta_1 = \{ v_1, pv_1, p^2v_1 .... p^{e1-1}v_1$	}	basis for $C_{T,v1} = J_{e1}(\lambda)$	e1 vectors
v_2	$\beta_2 = \{ v_2, pv_2, p^2v_2 .... p^{e2-1}v_2$	}	basis for $C_{T,v2} = J_{e2}(\lambda)$	e2 vectors
.....				
v_γ	etc. // γ is the dimension of Ker(p) = geometric multiplicity of eigenmanifold.			

The total number of vectors in the list above is the total number of dots in the dot diagram which is a, the algebraic multiplicity of the eigenmanifold. When we then do the above procedure for all eigenmanifolds of our matrix T, we then have a number of basis vectors which equals the dimension of the matrix T. These basis vectors are then gathered to make the columns of a square matrix P (same size as T), and then P is the similarity which brings T to J, the Jordan form.

Proof of Claim A: see page 68. From our previous theorem (with its flaky proof), we know that $p^{ei-1}v_i$ is a basis vector of Ker(p). Therefore, $p\,(p^{ei-1}v_i) = 0$, which says $p^{ei}v_i = 0 = p^{ei}(T)v_i$ in more detail. So certainly then $p^{ei}(T)$ is a *candidate* for whatever is the min poly associated with v_i which we call $m_{T,vi}$. However, the min poly might have a lower exponent than e_i. All we know is that $p^{ei}(T)v_i = 0$ and $m_{T,vi}(T)v_i = 0$, so as he says, all we know so far is that $m_{T,vi} \mid p^{ei}$. We know, however, that $p^{ei-1}v_i \neq 0$, since $p^{ei-1}v_i$ is a basis vector, so certainly the exponent of the min poly cannot be $e_i - 1$. Since we can write $p^{ei-1}v_i = p^{ei-1-n}\,p^nv_i$ for $n = 0,1,2...e_i - 1$, we cannot have $p^nv_i = 0$ for *any* of these powers, because if it were 0, then we would have that $p^{ei-1}v_i = 0$. Thus, the min poly exponent must in fact be e_i.

Proof of Claim B. On page 68 in item (a) we show that the various $C_{T,vi}$ spaces are contained in the space we want, which is $V_i = \ker(p^{pi})$. The way this is shown is simple: consider a generic element of the C space f v_i where f is any poly. We simply show here that $p^b(fv_i) = 0$. As before, we use the fact that $p^{ei}v_i = 0$. So in this way we show that all the C spaces are contained in the V_i space. If the all are separately, then the sum must be as well. Now the proof that the C spaces are independent is deferred to later on this page, and we assume that to show that the dims of the C spaces add up to $a = \upsilon(p^b)$, and this then clinches the fact that we have a direct sum of the C_i.

All this took about 1/2 page. The Lemma is the hard part! The Lemma shows that $p^{ej} \mid f_j$ and that therefore (as I show in pencil), if the sum $f_1v_1 + f_2v_2 + ... = 0$, then each term is separately 0, and this is how you prove independence.

The induction proof of the Lemma was too much for me at this moment. I see that it is shown to be true in the case that $e_1 = 1$. But then the second part attempts to show that if it is true for some $e_1 = N-1$, then it is also true for $e_1 = N$. I have trouble understanding what the Lemma says if we assume it is true for $e_1 = N-1$. Remember, this is just the first e_j in the chain of e's. As usual, I could plough through this thing and take another whole day, but I believe the result, so let's not do this. The proof of the lemma ends on page 70 where we start our "summary".

The Summary is exactly what I would write. The examples in the next section are VERY useful.

Some Jordan Form examples.

4.2.1 Example (a)

CASE 1. I accept that he has computed the char poly correctly and we have p_1 and p_2 to worry about, and here we are doing the p_1 part which is $p^2 = (x-2)^2$ where a = 2. He first computes p(A) and shows it as a 4x4 matrix, then he gets it to echelon form using ERO's. We see that rank = 2, so we know that our nullity $\upsilon(p^1) = 2$ = the geo mult of our main nullspace. But since $\upsilon(p^1) = 2 = a$, we know that b = 1 from his little trick -- where b is the min poly exponent. Since b = 1 and $\gamma = 2$, we know exactly how to make the dot diagram. Next, we examine the problem (A - 2I) [x,y,z,t] = [0,0,0,0] using the reduced form of the matrix (since ERO's don't change your eigenvectors), and we find that we need (A' - 2I)[x,y,z,t] = [x,z,0,0] so we need x=0 and z=0. So our eigenvector has the form [0,y,0,t] and we may then take the two eigenvectors to be [0,1,0,0] and [0,0,0,1]. These of course form a basis for Ker(p) = N(A-2I), just as he states. He then does the step I have not studied yet: he writes $\text{Ker}(p^1)$ = direct sum of two 1D cyclic spaces based on these little vectors v_1 and v_2 [I leave off the case 1 labels for now].

Comments added later: In this example, we are actually shown the matrix, and it is the full matrix T which is a 4x4, it is not just a submatrix in one eigenmanifold. Once we know that $\gamma = 2$, we know there are two boxes in the bottom row, but since a = 2 = total # boxes, we are now done! We then have two subspaces C each of which has dimension e = 1. The basis vectors found above for Ker(p) [those that satisfy pv = 0] are called v_{11} and v_{12} where the first index means we are doing CASE 1 now. So this eigenmanifold gives a simple $J_1 + J_1$ situation of two 1x1 matrices. These appear at the bottom of page 72 as the pair of 2's, since that is the eigenvalue.

CASE 2. Doing our same ERO's, we get rank = 3 so $\upsilon(p^1) = 4 - 3 = 1 = \gamma$. Since a = 2 for this case, we are not "done" yet in our search for b. Now we have to manually square the non-reduced matrix and using Maple I did this and found that the resulting matrix has 2 rows of zeros so has rank = 2 and $\upsilon(p^2) = 4-2 = 2$. You cannot do this with the "reduced" matrix, by the way. Now that we have $\upsilon(p^2) = 2 = a$, we now

know that b = 2. The dot diagram is easy to draw, b = 2, $\gamma = 1$. [By the way, we did not really have to compute p^2. We know that $b \leq 2$ and we saw that b = 1 did not "do the job". So we know right then that the correct value must be b = 2.]

Now we come to the question of finding a basis and here things are not so trivial. We are told that we should first find a basis for $\text{Ker}(p^2)$, a space we know has dimension a = 2. Here we see the matrix p^2 which I just computed in Maple (it agrees), and we see that it is rank = 2, he does the echelon ERO thing. We then do our usual cranking and we find that we need [x - 2z - t, y-10z-3t, 0, 0] = [0,0,0,0]. We solve for t and get one equation: y = 3x + 4z. Any vectors [x,y,z,t] which make this happy are OK, and we can take the two he shows. He picks them so that each one has a 0 somewhere, but they are not orthogonal, just basis vectors. So the basis vectors of the space $\text{Ker}(p^2)$ are these vectors X_1 and X_2.

What comes next? We now use our little "The Basis Business" paragraph above to claim that $<pX_1, pX_2>$ spans the space $N_{2,p}$. From our dot diagram, we know at once that dim $N_{2,p}$= 1, so we know that both these vectors $<pu_1, pu_2>$ cannot be LI. So he puts in p and computes these two column vectors. I checked the first one. Now having done this, he sees that they are linearly dependent, so we have to reduce to one vector. Now we knew that $\upsilon_{2,p} = \upsilon(p^2) - \upsilon(p^1) = 2 - 1 = 1$, so we are not surprised to find only one vector spanning our space $N_{2,p}$.

So, what we end up with in CASE 2 is a C space of dimension 2. The vector X_1 can be taken as the generating vector v_1, so we call it $v_{21} = X_1$. The two vectors which are a basis for this space are v_{21} and pv_{21} using the usual powers rule. He refers to this as the basis β_{21} since it is generated by this v_{21}.

The final act then is to make a P matrix using as columns the union of all our basis vectors. I did this and jammed it into Maple and it really works!

4.2.2 Example (b)

Here we are "given" a 6x6 matrix A, but he does not write down a specific matrix, he just gives us the properties of the matrix. We see that it has just one eigenmanifold since a = 6, and we see that b = 3. The eigenvalue of our one eigenmanifold is $\lambda = 0$, so this allows us to replace p by A, since p(A) = A. We then compute the three dimensions of the N spaces as shown, and this gives us at once the dot diagram. that is to say, we enter the rows, and then the columns are of course determined. We see at once that the Jordan block structure will be $J_3 + J_2 + J_1$, and this is shown in detail at the bottom of page 74.

Now the very interesting part of this "richer" example is how you compute the basis vectors which will be used in the similarity P. The algorithm presented here is that which I referred to above in Claim C (and which I never proved). Let's go through the steps:

(1) We start at the top row of the dot diagram and we consider $\text{Ker}(p^3)$. Since b = 3, we know this is a = 6, so we know this space has 6 basis vectors, and we name these X_1 ... X_6. We then use our "usual trick" to claim that the set of vectors $\{ p^2X_1, ... p^2X_6\}$ will span the hatch space $N_{3,p}$. Of course this space only has dimension 1, so only one of these 6 guys survives. We use the left-to-right idea and we pick the first non-zero one as our guy! Perhaps it is $p^2 X_4$. (remember that $p^n = A^n$ in this example!). So whatever this X_4 comes out to be, call it v_{11}. Then our sole basis vector for $N_{3,p}$ is A^2v_{11}.

(2) Now we move one row down in the dot diagram. In general, we move down to the next row where the width has increased, because only then do we arrive at an N space that is larger and different. But here we step out one on each row, so each row presents a new N space. So we go down to the row for $N_{2,p}$. As before, we think first about basis elements for the space $\text{Ker}(p^2)$, one lower power than in (1) above. We

know that $\upsilon(p^2) = 5$, so there will be 5 basis vectors which we call $\{ Y_1....Y_5 \}$. We then move to $N_{2,p}$ in the usual way: $\{ p^1Y_1, ...p^1Y_5 \}$. But $\upsilon_{2,p}= 2$, so only 2 of these can be LI.

At this point we do a critical thing. Since our space $N_{2,p}$ is really just an "enlargement" of $N_{3,p}$, we can use the basis vectors (only one in this case) of $N_{3,p}$ as part of the basis of the larger space $N_{2,p}$. So we add our previously found basis vector A^2v_{11} to the start of the list as shown. Then we scan the rest of the list and find the first vector that is non-zero and which is not a multiple of the first vector. We call this vector Av_{12}. The idea is that we are "extending" the basis of $N_{3,p}$ to get a basis for $N_{2,p}$.

(3) Now we drop down yet another row to the bottom and last row, and do the above process once more. We first find a basis for $Ker(p^1)$ which is Ker(p) and has dimension $\gamma = 3$. We assume the basis is the set of vectors $\{ Z_1, Z_2, Z_3 \}$. We then increase our list by adding the previously found basis functions to get this list: $\{ A^2v_{11}, Av_{12}, Z_1, Z_2, Z_3 \}$. Again, we are doing an "extension" of the basis of $N_{2,p}$ to get a basis for $N_{1,p}$. We then pick the first non-zero of the Z's which is LI of the first two vectors, and we call it v_{13}.

So, at this point, we have the following basis vectors for the bottom row: $\{ A^2v_{11}, Av_{12}, v_{13} \}$. These are the genuine-issue eigenvectors of our matrix! There are 3 because γ = geo mult = 3. These are the basis vectors for Ker(p). Notice carefully please that Av_{11} and v_{11} are not in this set, only A^2v_{11}.

Now, we have our three C spaces of dimensions 1,2 and 3 which add up to the alge mult of 6 and we have our final J matrix as shown page 74 end. We then create the P matrix by listing off all the basis vectors of the three C spaces. These are as listed center of page 74. The convention seems to be to put the vector first, then the higher applied powers of p, this then is the standard Jordan cyclic basis ordering.

Since this example did not have a detailed matrix showing, I could not use Maple to verify that this P does in fact give the J matrix shown, but I do believe it!

Teach by example!

At this point, I printed out and perused the remaining pages of Chapter 4 of these notes. Unlike earlier chapters, these notes are in several different PDF files which I have tried to number in some reasonable manner. I have not in any way "studied" these sections, but in perusing them, I found many interesting facts which I might make use of at some later time, so here are a few notes.

4.3 Uniqueness of the Jordan Form.

If you agree to put the Jordan Blocks in order of decreasing size within each eigenmanifold, then the only issue of non-uniqueness is the order in which you put your eigenmanifolds. Since eigenvalues might be complex, you can't just put them in decreasing order.

So OK, this is not an exciting topic to me, I know it is true. Along the way, author happens to mention a few buzzwords relating to the dot diagram: the set of column heights $\{e_i\}$ is the Segre characteristic, and the set of row widths $\{ \upsilon_{h,p}\}$ is the Weyr characteristic.

On page 77, he shows to different Jordan forms which have the same a and b, so these numbers are not sufficient to determine the Jordan form. In this case we could distinguish the two cases he shows by having geo mult = 2 and 3.

4.4 Non-derogatory matrices.

This is a matrix which has $b_i = a_i$ for all the p factors, so char poly = min poly. Well, in each eigenmanifold, you have a boxes but the column height is b = a, so it must be that you have a single column for your dot diagram. This means that each eigenmanifold is just one big Jordan block whose size is the alge mult a. Not sure where the name comes from, nor are we given a significance of this special case. The OED says nothing relevant about the word derogatory other than it implies "lowering", so you might say that the complexity of the Jordan blocks are not "lowered" but have their "full form".

4.5 Calculating powers of J and A.

Author gives us explicit banded matrices for arbitrary powers of a J matrix. More power means more off-diagonal activity. Because the eigenvalue appears in the form λ^m , if $\lambda < 1$ then as m increases, the matrix tends toward 0. This fact is stated on page 80. Then on page 81 we can say this: if matrix A has all eigenvalues < 1 in magnitude (even if complex) then $A^m \to 0$ as $m \to \infty$. You just use the fact that the J matrix does this to prove this fact. Not sure where we might use this.

4.6 The notion of the matrix e^A

Formally we know how to write this as a power series. To prove it, we have to show convergence, and this is quickly shown in the proof page 82.

4.7 Properties of exponential matrices.

This is a great collection, I hope I can remember it is here. Each is proven. They all look familiar to me.

Number (vii) says that $e^A = p(A)$, a polynomial in A. The reason is that we know that A^n eventually reaches its max power and can then be expanded in lower powers, this is the min poly of A. We know that the degree of such a poly must be n^2 or less, and in fact we know it is less than degree n which is the order of A, because the char poly is of this order in x. We know from C-H that ch(A) = 0 and ch is degree n. In any event, the polynomial p(A) is different for different matrices A.

Number (x) gives an exponentiated J block.

4.8 Systems of Differential Equations.

Here we look at a super simple first order dX/dt = AX system. We end up at once with the solution $X = e^{tA} X_0$ and at once we see the usefulness of knowing something about exponentiated matrices!

I am not sure I have ever seen this before. I don't offhand see any book I have that shows this simple thing!

Now what can do to simplify the solution to this problem so we are not stuck with $X(t) = e^{tA} X_0$? The answer is that you use your Jordan form! On page 87 he gives a nice example. We have a 3x3 matrix A, and details are not given, but he gives the Jordan decomposition without telling us what P is.

Aside: how does Maple do this? Yes! Just say jordan(A, 'P'). So I will enter the matrix A and tell it to compute J and P. Ouch, Matthews has made a goof here. The matrix A which he shows does not have the eigenvalues he claims, so probably some sign or number is wrong. But OK, let's assume some other matrix works as he says. Then he shows how you can eliminate the expo matrix from the problem.

There are probably simpler ways to solve such a system, but this is elegant.

4.9 Markov Matrices

No motivation is given here, but I think these describe finite-state Markov chains somehow, where the entries in the matrix are probabilities of moving between states (a transition matrix). The matrix he talks about here has rows which add up to 1, and no negative matrix elements. He then proves some interesting facts about such matrices:

product of two Markovs is Markov

eigenvalues are all ≤ 1 in magnitude

if all matrix elements are strictly positive (you write $A > 0$) then 1 is an eigenvalue and all other eigenvalues are < 1 in magnitude, and there is only one eigenvector for the $\lambda = 1$ eigenvalue.

If you raise a positive Markov matrix to large power m, eventually you get a stable form where all the rows are the same and are equal to the eigenvector (transposed) that goes with the 1. Page 92 gives an example which shows how this happens.

So this is just an oddball little section he has thrown in .

4.10. The Real Jordan Form.

Just go right to the example on page 98. We have a 4x4 matrix A. If we take this to the usual Jordan form, we have to deal with complex eigenvalues. So the Jordan form will have complex numbers on the diagonal. If you don't want complex numbers, there is a "form" you can get to which is less diagonal than the Jordan form, but it has the advantage that all numbers are real. For this sample problem, the "real Jordan form" is shown on page 99 or page 97. On the diagonal you have little 2x2 blocks as shown which are somehow "replacing" the complex eigenvalues, and the off diagonal has 2x2 identity matrices. The general idea is shown on page 94. I think this K thing is what replaces the Jordan block. In any event, you get a banded matrix that has a "band width" of 4 instead of 2 as you get with the complex Jordan form.

He does this same example again on page 102 using a method (from some other person) which also gives you the P matrix that does the job. The results for this example as then shown on page 104.

So fine, I am happy to know that this "exists". I am reminded of the trick in the QR algorithm which also gets rid of complex numbers in a numerical method.