VIGRE Algebra QR Review, October 21, 2002

Similarity of matrices: Jordan and rational form

Fix a positive integer n. We consider only $n \times n$ matrices over a field k. $M_n(k)$ denotes the ring of $n \times n$ matrices over the field k. This is an associative ring with identity. It is commutative iff $n = 1$. The invertible $n \times n$ matrices form a group under multiplication, the group of units in the ring $M_n(k)$. This group is denoted $\mathrm{GL}(n, K)$. If V denotes an n-dimensional vector space over k, one gets a k-linear isomorphism of $\mathrm{End}_k(V) \cong M_n(k)$ by fixing an ordered basis $\mathcal{B} \ = v_1, \ldots, v_n$ for V over k. The k-linear endomorphism $T : V \to V$ corresponds to the matrix whose j th column is the value of $T(v_j)$ written in coordinates with respect to the same basis $\mathcal{B}$. If V is k^n with the standard basis (often denoted $e_1, \ldots, e_n$), and A is the matrix of T with respect to the standard basis, then the value of T on w, where w is an element of k^n written as a column vector, is Aw, also written as a column vector.

If $\mathcal{B}'$ and $\mathcal{B}$ are two bases for V, and B is the matrix whose j th column consists of the coefficients needed to write v_j in $\mathcal{B}$ in terms of $\mathcal{B}'$, and the matrix of T with respect to $\mathcal{B}$ is A, then the matrix of T with respect to $\mathcal{B}'$ is BAB^{-1}. Multiplying by B^{-1} changes $\mathcal{B}'$-coordinates to $\mathcal{B}$-coordinates, multiplying by A gives the value of T in $\mathcal{B}$-coordinates, and multiplying by B changes back to $\mathcal{B}'$-coordinates. BAB^{-1} and A represent the same linear transformation and are called *similar* or *conjugate* matrices.

In fact, in any ring R with identity, if u is a unit of R then the map θ_u that sends r to uru^{-1} is a ring automorphism of R, and the map that sends u to θ_u is a group homomorphism from the multiplicative group of units of R into the group of ring automorphisms of R. Conjugate elements of a ring are indistinguishable in terms of their ring-theoretic properties, e.g., whether they are nilpotent, whether they are idempotent (square equal to the identity) or have order k in the multiplicative group of the ring, etc.

Let x be a variable over k. For each polynomial $f(x) \in k[x]$, say $c_n x^n + \cdots + c_0$, where the $c_j \in k$. We define $f(A)$ to be $c_n A^n + \cdots + c_0 I_n$, where I_n is the $n \times n$ identity matrix. Thus, $f(A)$ is again a matrix. We can make k^n into a $k[x]$-module by the rule $f(x)v = f(A)(v)$. It turns out that A and A' are similar if and only if the $k[x]$-modules they define are isomorphic.

A matrix is invertible if and only if its determinant is nonzero. A matrix is not invertible if and only if it has a non-trivial kernel. If $\lambda \in k$ is a scalar, there is a nonzero vector v such that $Av = \lambda v$ iff $\lambda I - A$ has a non-trivial kernel iff the determinant of $\lambda I - A$ vanishes iff λ is a root of the polynomial $\det(xI - A)$, which is a monic (leading coefficient 1) polynomial in x. This polynomial is called the *characteristic polynomial* of A, and its roots are called the *eigenvalues* or *characteristic values* or *characteristic roots* of A. Note that in algebraic closure $\overline{k}$ of k, there are always n eigenvalues, counting multiplicities.

Over an algebraically closed field, every matrix is similar to an upper (resp., lower) triangular matrix with the eigenvalues on the diagonal. (This is a weak version of Jordan form, and easy to prove. Choose an eigenvector and make it first in a basis. The matrix will have all entries zero in the first column below the top entry. Use induction, working with the $n-1$ size square matrix in the lower right corner. To get a lower triangular matrix put the first eigenvector you find last in the basis, and then apply induction to the size $n-1$ matrix in the upper left corner.)

The characteristic polynomials of similar matrices are the same. If f is the characteristic polynomial of A, then $f(A) = 0$ (the Cayley-Hamilton theorem). The least degree polynomial g such that $g(A) = 0$ is called the *minimal* polynomial of A. The polynomials that vanish on A form an ideal in $k[x]$: the minimal polynomial is the monic generator, and the characteristic polynomial is in this ideal and so is divisible by the minimal polynomial. Matrices in the same similarity class have the same minimal polynomial. If h is any polynomial, the eigenvalues of $h(A)$ can be obtained, with correct multiplicities, as the values of $h(\lambda)$ as a λ runs through all eigenvalues of A (to see this, work with an upper triangular matrix similar to A). Every eigenvalue of A occurs as a root of the minimal polynomial of A: the characteristic and minimal polynomials have the same roots, but with possibly different multiplicities. A matrix A is nilpotent if and only if all eigenvalues are 0. In this case, the characteristic polynomial is x^n, and $A^n = 0$.

If A is invertible, then AB and BA are similar, since $AB = A(BA)A^{-1}$. Similarly, if B is invertible we have $BA = B(AB)B^{-1}$. Thus, when A is invertible, AB and BA have the same characteristic polynomial. However, this is true for all matrices, without invertibility. The reason is that it suffices to prove it for matrices of mutually distinct indeterminates over $\mathbb{Z}$, and such matrices are invertible over the field generated by their entries over $\mathbb{Q}$.

Let $f = f(x) \in k[x]$ be a monic polynomial of degree $n \geq 1$. We can write f as $x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, where the $a_i \in k$. Consider the cyclic module $V = k[x]/(f)$, which, as a vector space over k, has basis $1, x, x^2, \ldots, x^{n-1}$. Multiplication by x is a k-linear transformation T on V. Note that $xv_i = x \cdot x^{i-1} = x^i = v_{i+1}$ for $i < n$ while $xv_n = -a_0 v_1 - \cdots - a_{n-1}v_{n-1}$. The matrix of T with respect to multiplication by x (some authors use the transpose instead) is called the *companion matrix* of f. It has a string of 1's just below the main diagonal. The last column consists of the negatives of the coefficients of f (other than the leading coefficient). All other entries are 0. The characteristic and minimal polynomials of this matrix are both f. When this matrix is used to create a module, it gives the cyclic module $k[x]/(f)$.

The negative of the coefficient of x^{n-1} in the characteristic polynomial of A is the same as the sum of the diagonal entries, and is called the *trace* of A. Similar matrices have the same trace. Note that because AB and BA have the same characteristic polynomial, they have the same trace. This implies that the trace of $AB - BA$ is always 0. Note also that if f is the characteristic polynomial of A, the constant term is $\det(0 \cdot I - A) = \det(-A) = (-1)^n \det(A)$.

The direct sum of square matrices $A_1, \ldots, A_h$, where A_j has size d_j, is defined to be the square matrix of size $d_1 + \cdots + d_h$ which, in block form, has $A_1, \ldots, A_h$ strung along the main diagonal, while all off diagonal blocks are 0. E.g., in block form the direct sum of A_1 and A_2 is $\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$. The 0 in the first row is a $d_1 \times d_2$ matrix all of whose entries are 0, while the 0 matrix in the second row is $d_2 \times d_1$. The characteristic polynomial of a direct sum is the product of the characteristic polynomials of the individual A_i, while the minimal polynomial is the least common multiple of the minimal polynomials of the A_i.

Rational (canonical) form for a matrix. Given A, we want to single out a matrix in its similarity class that is unique. We can use the structure theorem for modules over a PID to do this. The $k[x]$-module determined by A must be a torsion module. The structure theorem implies that it is a direct sum of cyclic modules $k[x]/(f_1) \oplus k[x]/(f_2) \oplus \cdots \oplus k[x]/(f_h)$, where the f_j are monic polynomials of degree > 0 such that each divides the next (i.e., $f_1|f_2, \ldots, f_{h-1}|f_h$). This means that A is similar to a matrix which is the direct sum of h companion matrices A_i, where A_i is the companion matrix of f_i. For each block A_i, the characteristic and minimal polynomials of A_i are both f_i. The divisibility condition on the f_i guarantees uniqueness: there is one and only one matrix similar to A that is in rational form: notice that the entries are in the same field k as the entries of A. As a corollary, if two matrices with entries in a field k are similar over a larger field k', then they are similar over k: they will have the same rational form over k', and therefore over k, and each will be similar over k to its rational form over k. (For some authors, the rational form will be the transpose of what we have defined here.) Note that the minimal polynomial of A is f_h and the characteristic polynomial is the product $f_1 \cdots f_h$.

Jordan (canonical) form for a matrix. Now assume that the field is algebraically closed. The module corresponding to the matrix A can be written as a direct sum of cyclic modules which are primary, each having the form $k[x]/(x-\lambda)^d$, where $\lambda \in k$ (over an algebraically closed field, all higher degree polynomials factor into linear factors, which are the only irreducibles). Now, when the cyclic module is $k[x]/(x-\lambda)^d$, if B is the size d matrix that gave rise to it, then $B - \lambda I$ is nilpotent, and its rational form has a string of entries equal to 1 just below the main diagonal and 0's elsewhere. This means that B is similar to a matrix with all diagonal entries on the main diagonal equal to λ and a string of 1's just below the diagonal, with all other entries 0. Such a matrix is called a *Jordan block*. Every matrix is similar to a direct sum of Jordan blocks, and this is unique up to the order in which the blocks appear. One can arrange the blocks so that all blocks involving a certain eigenvalue λ are consecutive, and sizes for blocks involving λ increase as one moves down the diagonal. Some authors describe Jordan form as the transpose of what we have used here. (Exercise: show that every matrix over a field is similar to its transpose.) In the minimal polynomial, the multiplicity of an eigenvalue will be the size of the largest Jordan block in which it occurs, while in the characteristic polynomial the multiplicity of an eigenvalue is the sum of the sizes of the blocks in which it occurs.

Suppose that x is to act on V the way the matrix A does. One can understand what the module is as follows. Consider the matrix $xI - A$ over $k[x]$. If one kills its column space one gets precisely a module such that x acts the way A does. One can do elementary row and column operations on the matrix $xI - A$ until it is diagonal. If the polynomials f_i occur on the diagonal, then the module is the direct sum of the $k[x]/(f_i)$.

A matrix is called *cyclic* if the corresponding module is cyclic. A matrix is cyclic if and only if its characteristic and minimal polynomials are the same. A matrix is called *diagonalizable* if it is similar to a diagonal matrix. A matrix is diagonalizable if and only if its minimal polynomial is square-free, i.e., has all distinct roots. A sufficient (but not necessary) condition for a matrix to be diagonalizable is that it have n distinct eigenvalues: since each must occur in a Jordan block, and the sum of the block sizes is n, each must by occur with multiplicity 1 even in the characteristic polynomial, which in this case will turn out to be the same as the minimal polynomial.

Problems new this week

1. Let T be a linear transformation of a vector space T such that T has minimal polynomial $x^2(x-1)^2$ and characteristic polynomial $x^3(x-1)^4$. What are the possible Jordan canonical forms (resp., rational canonical forms for (the matrix of) T?

2. (a) Find the Jordan canonical form of the complex matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$.

(b) Let F be any field. Suppose that a normal subgroup N of $\mathrm{GL}(2, F)$ contains one matrix of the form $\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ with $a \in F - \{0\}$. Prove that N contains every matrix of this form.

3. Let A and B be two $n \times n$ matrices with complex entries.

(a) Show that for $\lambda \in \mathbb{C}$, λ is an eigenvalue of AB if and only if λ is an eigenvalue of BA.

(b) Show that if A or B is invertible, then AB and BA are conjugate.

(c) Give an example for $n = 2$ where AB and BA are not conjugate.

4. Let A be an $n \times n$ matrix over a field k of characteristic 0, and suppose that for every positive integer t, the matrix A^t has trace 0. Prove that A is nilpotent.

5. Let M be the ring of all $n \times n$ matrices over a field. The group G of invertible matrices acts on M by conjugation. Show that a conjugate of a nilpotent matrix is nilpotent. When $n = 6$ count the number of conjugacy classes of nilpotent matrices, and justify your answer.

6. Let A, B be two square $n \times n$ matrices satisfying over a field of characteristic 0 satisfying $AB - BA = A$. Show that $A^kB - BA^k = kA^k$ for every positive integer k. Deduce from this that A is a nilpotent matrix. (The char. 0 hypothesis was omitted on the original exam. As an extra problem, show that this is false in char. 2 for $n = 4$.)

7. (Not a QR problem.) Let T be a linear transformation over $\mathbb{R}$ on a finite-dimensional vector space V. Show that there is either a one-dimensional or a two-dimensional subspace of V that is mapped into itself by T. (One suggestion: over $\mathbb{C}$, T has at least one eigenvalue, and a corresponding eigenvector.)