8000 Fall 2006 Day 3.
Canonical forms of matrices, the power of Cramer's rule

Our decomposition theorem gives us a standard model in each isomorphism class of finitely generated torsion k[X] modules. This will be used next to provide a standard matrix representative for each conjugacy class, or similarity class as it is usually called, in the ring $Mat_n(k)$, of n by n matrices over any field k.

Recall that a linear map T:V--->V on a k vector space V, provides a unique k algebra map k[t]--->$End_k(V)$, sending t to T, and hence f(t) to f(T), and hence a unique k[t] module structure on V. We will denote $End_k(V)$ simply by End(V) in this chapter for brevity, since we will not be concerned with the larger ring of group endomorphisms.

Conversely, a k[t] module structure on V singles out a unique linear map T, the image of t under the map k[t]--->$End_k(V)$. Thus k[t] module structures on V are in natural bijection with the elements of End(V). We want to ask what equivalence relation is imposed in this way on End(V) by considering isomorphism classes of modules.

Note that if f:(V,T)--->(V,S) is a k[t] module isomorphism, then f is a k isomorphism that takes multiplication by (i.e. application of) T into multiplication by S. Thus f(Tv) = S(fv) for every v in V. Since f is an isomorphism this implies Tv = $(f^{-1}Sf)v$, for every v. Hence S and T are conjugate by the isomorphism f.

Conversely, these equations show that if T = $(f^{-1}Sf)$, then T and S define isomorphic k[t] modules via the isomorphism f. Thus isomorphism classes of k[t] module structures on V correspond to conjugacy classes of endomorhisms via the action of Aut(V) on End(V).

Hence when V has finite k dimension, our canonical models of each k[t] - isomorphism class, translate into canonical representatives of each conjugacy class in End(V). Recall each finitely generated torsion k[t] module (V,T) has a model $V \approx k[t]/f_1 \times .... \times k[t]/f_m$, where each f_i is a monic polynomial in k[t], and f_i divides f_{i+1}.

Under the isomorphism $(V,T) \approx k[t]/f_1 \times .... \times k[t]/f_m$ the linear map T:V--->V, i.e. multiplication by T, becomes multiplication by the variable t on each factor of $k[t]/f_1 \times .... \times k[t]/f_m$. Hence if we choose a natural k basis for this model vector space, the resulting matrix for t will give a natural matrix representing T in some corresponding k basis for V.

A k - basis for $k[t]/f_1 \times .... \times k[t]/f_m$, can be obtained as the union of bases for each factor space $k[t]/f_i$, and the simplest basis for $k[t]/f_i$, is $\{1, t, t^2,..., t^{r_i-1}\}$, where f_i has degree r_i. If $f = a_0 + a_1t + ...+ a_{r-1}t^{r-1} + t^r$,

the matrix of t in this basis is this: $\begin{bmatrix} 0 & 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & 0 & -a_2 \\ .. & .. & .. & .. & .. \\ 0 & 0 & .. & 1 & -a_r \end{bmatrix}$, where the jth column

is the coefficient vector of t times the jth basis vector. E.g. $t(1) = 0(1)+1(t) + 0(t^2) + ...+ 0(t^{r-1})$, gives the first column.

This is called a cyclic basis, since the linear map carries each basis vector to the next one, except for the last one, which is carried to a linear combination of the basis by means of scalars which are precisely minus the coefficients of the polynomial f. This is called a companion matrix C_f for f. [Other versions of it in other books may have the coefficients of f along the bottom, and the 1's above the diagonal.] Note that if $v_1,...,v_n$ is one cyclic basis for (V,T) then for any $c \neq 0$, $cv_1,...,cv_n$ is another, so cyclic bases are never unique.

If $f_1,...,f_m$ is the sequence of polynomials defining the module (V,T), the full matrix for T using the cyclic bases for each factor looks like this:

$\begin{bmatrix} \begin{bmatrix} C_{f_1} \end{bmatrix} & & & \\ & \begin{bmatrix} C_{f_2} \end{bmatrix} & & \\ & & \cdots & \\ & & & \begin{bmatrix} C_{f_m} \end{bmatrix} \end{bmatrix}$, where there are zeroes away from the C_{fi}.

Summarizing, we have the following.

Theorem: If V is a vector space of finite dimension n over a field k, and T is any linear endomorphism of V, there exist bases for V in which the matrix of T is composed of one or more blocks, each block being a companion matrix for a monic k polynomial f_i.

The sum of the degrees of the f_i equals n, and we may choose them so each f_i divides f_{i+1}. If we do this, then two maps S,T of V are conjugate if and only if they have exactly the same matrix of companion blocks. There is exactly one companion matrix block C_f for each factor k[t]/(f) in the standard decomposition of the k[t] module structure for (V,T). Each block C_f has dimension deg(f) by deg(f).

Terminology: We call the unique matrix of this type associated to T, the <u>rational canonical matrix</u> for T.

Two natural questions remain:

1) how do we find the canonical form for a given matrix? and (more difficult):

2) how do we find a basis that puts a given matrix into canonical form?

A third question is:

3) is there a simpler canonical matrix in cases where the polynomials f_i are particularly simple,

e.g. when they all factor into linear factors over k?

Before addressing these questions, we derive some useful consequences of the results we already have. For example we can already compute the important invariant $\prod f_i$ of the module (V,T), using determinants. Briefly, we claim this product is the "characteristic polynomial" of T, $\prod f_i = \det[tI-T] = ch_T(t)$. Since f_m is the annihilator of the module (V,T), this implies the Cayley Hamilton theorem: $ch_T(T) = 0$.

Before proving this, we recall the basic theory of determinants, including LaGrange's formulas for expanding them along any row or column, and the resulting "Cramer's rule".

Review of determinants.

If $A = [a_{ij}]$ is an n by n matrix over a commutative ring, denote by A_{ij} the (n-1) by (n-1) matrix obtained from A by deleting the ith row and jth column. Then LaGrange's formulas say, for each fixed value of i, $\det(A) = \sum_j (-1)^{i+j} \det(A_{ij})$, (expansion by the ith row), and for each fixed value of j, $\det(A) = \sum_i (-1)^{i+j} \det(A_{ij})$, (expansion by the jth column.

Thus if we define adj(A) = the adjoint of A, as the matrix whose i,j entry equals $(-1)^{i+j} \det(A_{ji})$, i.e. as the transpose of the matrix of signed determinants of the A_{ij}, it follows that the matrix products adj(A).A = A.adj(A), both equal the diagonal matrix det(A).I, whose entries along the diagonal are all equal to det(A).

Thus if det(A) is a unit in the ring of coefficients, then A is an invertible matrix with inverse equal to $(\det(A))^{-1}$.adj(A). Since for any two n by n matrices A,mB we always have det(AB) = det(A)det(B), the converse is also true. I.e. AB = I implies det(A)det(B) = det(I) = 1, so both det(A) and det(B) are units. Thus the equation adj(A).A = A.adj(A) = det(A).I, yields a formula for the inverse of an invertible A, and hence Cramer's rule for solving invertible systems AX=Y.

Cramer's formula also implies that a matrix and its transpose have the same determinant. I.e. since the transpose of the adjoint is the adjoint of the transpose, taking the transpose of the equation adj(A).A = A.adj(A) = det(A).I, gives $(\det(A^t).I) = A^t.adj(A^t) = adj(A^t).A^t = (\det(A).I)^t = \det(A).I$, the last because the diagonal matrix det(A).I is symmetric.

Define: the **characteristic polynomial** of a linear map T on a finite dimensional space $ch_T(t) = \det([tI-A])$ where A is any matrix for T.

By the previous remarks, a matrix A and its transpose A^t have the same characteristic polynomial.

Note: If A,B are two matrices matrix for T, A and B are conjugate, i.e. $B = C^{-1}AC$ for some invertible C. Then since $\det(B) = \det(C^{-1}AC) = \det(C^{-1})\det(A)\det(C) = \det(A)\det(C^{-1})\det(C) = \det(A)$, we see A and B have the same determinant. Similarly, [tI-A] and $C^{-1}[tI-A]C = [C^{-1}tIC-C^{-1}AC] = [tI-B]$ have the same determinant, since t.I commutes with every matrix C. Hence the characteristic polynomial of T is well defined by any matrix for T. It is easy to see the constant term of $ch_A(t)$ is $\pm \det(A)$, and the coefficient of t^{n-1} is minus the trace of A, (minus

the sum of the diagonal entries).

Exercise: If C_f is a companion matrix for the monic polynomial f, then $ch(C_f) = f$. [hint: use induction and expand across the first row.] One sees immediately the trace of C_f is $-a_{n-1}$.

Corollary:(Cayley Hamilton) If T is any linear transformation, then $ch_T(T) = 0$. In particular a matrix satisfies its characteristic polynomial.
proof: The annihilator ideal of the cyclic module R/I where I is any ideal of the ring R, equals I. In particular the annihilator ideal of $k[t]/(f)$ is (f). Hence the annihilator of the module $k[t]/f_1 \times ....x\ k[t]/f_m$, where f_i divides f_{i+1}, is f_m. I.e. the smallest degree monic polynomial f such that $f(t) = 0$ on this module is f_m. If this module represents (V,T), then the minimal polynomial of T is f_m, and we just showed the characteristic polynomial of T is the product $\prod f_i$. So the minimal polynomial of T divides its characteristic polynomial, which implies the corollary. **QED.**

Note: Since every factor f_i divides f_m, this proof shows that every irreducible factor of $ch_T(t)$ is an irreducible factor of the minimal polynomial $m_T(t)$, (and vice versa). Moreover, for a cyclic or companion matrix, the minimal and characteristic polynomials are equal. This is the analog of the fact that for a cyclic group Z/nZ, the order n of the group equals the annihilator of the group.

Example: A nilpotent matrix A is a square matrix such that $A^m = 0$ for some m. If A is nilpotent, follows that $A^n = 0$, where n is the dimension of the matrix A. Since all coefficients a_i of the characteristic polynomial for a nilpotent matrix are 0 except the leading one, the rational canonical form of a nilpotent matrix consists of blocks of form:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & .. & 1 & 0 \end{bmatrix}$$

The reader should verify this matrix is nilpotent.

Direct proof of Cayley Hamilton:

Cramer's rule implies the Cayley Hamilton theorem directly, without using the decomposition theorem, or the rational canonical form, as follows. Let [tI-A] be the characteristic matrix for A, with coefficients in k[t], and substitute t = A into this matrix, obtaining an n by n matrix with coefficients in the subring k[A], of $Mat_n(k)$.

This may be viewed as defining a linear map on the product space $(k^n) \times ....x\ (k^n)$, a product of n copies of k^n. Note this is not the same as substituting t = A into tI-A viewed as a polynomial with matrix coefficients, as that would give A.I-A = 0. Our result instead is the following n by n matrix M:

$M = \begin{bmatrix} A-a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & A-a_{22} & \dots & -a_{2n} \\ \dots & \dots & \dots & \dots \\ -a_{n1} & -a_{n2} & \dots & A-a_{nn} \end{bmatrix}$. Now take the transpose of this,

$M^t = \begin{bmatrix} A-a_{11} & -a_{21} & \dots & -a_{n1} \\ -a_{12} & A-a_{22} & \dots & -a_{n2} \\ \dots & \dots & \dots & \dots \\ -a_{1n} & -a_{2n} & \dots & A-a_{nn} \end{bmatrix}$, and apply it to the column of vectors $\begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$

in $(k^n)^n$.

By definition of the entries in A, this yields $M^t \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$. Now multiply

M^t from the left by $\mathrm{adj}(M^t) = (\mathrm{adj}(M))^t$. By Cramer's rule $\mathrm{adj}(M^t)\, M^t =$

$\mathrm{ch}_{(A^t)}(A).I = \mathrm{ch}_A(A).I =$ annihilates the vector $\begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$. I.e. the matrix

product $\begin{bmatrix} ch_A(A) & 0 & \dots & 0 \\ 0 & ch_A(A) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & ch_A(A) \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$. Hence $\mathrm{ch}_A(A)(e_i) = 0$ for

each i, so $\mathrm{ch}_A(A) = 0$. **QED.**

Note: This proves the minimal polynomial divides the characteristic polynomial, but does not show they have the same irreducible factors.

The canonical presentation of (k^n, A) by the characteristic matrix of A.

Next ask how to find the rational canonical form of a given n by n matrix A over a field k. Since it is determined by the cyclic decomposition of the k[t] module (k^n, A), it suffices to

diagonalize any presentation matrix for this module. So we look for a matrix M of polynomials in k[t], whose cokernel is isomorphic to (k^n, A) as k[t]- modules. Perhaps not surprisingly, it is given by the only k[t] matrix we know, the characteristic matrix [tI-A].

It is easy to find an explicit sequence of k[t] generators for (k^n,A), since $e_1,...,e_n$ are k generators, hence also k[t] generators of k^n. The map $(k[t])^n$--->k^n, sending E_i to e_i, where E_1 = (1,0,...,0) in $(k[t])^n$, and e_1 = (1,0...,0) in k^n, is thus a surjective k[t] module map, where $\sum f_i(t)E_i$ in $(k[t])^n$ goes to $\sum f_i(A)e_i$ in k^n.

The next theorem is our main result.

Theorem: Given an n by n matrix A over a field k, defining a k[t] module structure on k^n, the k[t] module map $(k[t])^n$--->k^n, sending $\sum f_i(t)E_i$ to $\sum f_i(A)e_i$, is surjective. Its kernel is a free k[t] module of rank n generated by the columns of [tI-A], the characteristic matrix of A. I.e. the following sequence of k[t] modules is exact: 0--->$(k[t])^n$--->$(k[t])^n$--->k^n--->0, where the left map is multiplication by [t.I-A].

Remark: This will follow from a version of the wonderful "root factor" theorem, but first we work an example using it.

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. Then det(A) = 0, since the sum of the 1st and 3rd rows is twice the middle row. The trace is 15, so we know that $ch_A(t) = t^3-15t^2+?t$. Dr Shifrin calls the unknown coefficient Fred(A). So we only need to compute Fred. A tedious calculation with polynomials and lots of minus signs reveals that the characteristic polynomial $\det(tI-A) = t^3-15t^2-18t$, so Fred is -18. But we will compute this another way.

To obtain the minimal polynomial of A, we use the previous theorem which says [tI-A] is a presentation matrix for the k[t] module (k^n,A), so we want to diagonalize [tI-A], i.e.:

$\begin{bmatrix} t-1 & -2 & -3 \\ -4 & t-5 & -6 \\ -7 & -8 & t-9 \end{bmatrix}$. Non zero constants are units, so we switch the first two columns, and multiply the 2nd row by 2, getting:

$\begin{bmatrix} -2 & t-1 & -3 \\ 2(t-5) & -8 & -12 \\ -8 & -7 & t-9 \end{bmatrix}$. Now add (t-5) times the 1st row to the 2nd row, and -4 times the 1st row to the 3rd row, getting:

$$\begin{bmatrix} -2 & t-1 & -3 \\ 0 & t^2-6t-3 & 3-3t \\ 0 & -3-4t & t+3 \end{bmatrix}$$, which immediately gives $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & t^2-6t-3 & 3-3t \\ 0 & -3-4t & t+3 \end{bmatrix}$$,

(think about it). Add 3 times the 3rd row to the 2nd, and switch 2nd and 3rd columns:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 12 & t^2-18t-12 \\ 0 & t+3 & -3-4t \end{bmatrix}$$, multiply the 3rd row by -12, and add to it (t+3) times the 2nd, yielding:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 12 & t^2-18t-12 \\ 0 & 0 & t^3-15t^2-18t \end{bmatrix}$$, hence $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & t^3-15t^2-18t \end{bmatrix}$$.

This shows our k[t] module is cyclic, isomorphic to $k[t]/(t^3-15t^2-18t)$, so t^3-15t^2-18t, is both minimal and characteristic polynomial of A.

The rational canonical matrix for A (and every matrix conjugate to A) is:

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 18 \\ 0 & 1 & 15 \end{bmatrix}$$. Please check me on this, as I am pretty weak at computation.

As corollary of the theorem above we get another proof of

Cayley Hamilton: If the k[t] module (k^n, A) is isomorphic to the product $(k[t]/f_1) \times \ldots \times (k[t]/f_m)$, in standard form, i.e. where f_i divides f_{i+1}, then the minimal polynomial of A is f_m and the characteristic polynomial is the product $\prod f_i$.

proof: Since [tI-A] is a presentation matrix for this module, there exist invertible matrices A, B over k[t] such that A[tI-A]B is diagonal, with lower diagonal entries equal to the f_i, and higher diagonal entries = 1.

Hence $\det(A)ch_A(t)\det(B) = \prod f_i$. Since A, B are invertible over k[t], their determinants are units in k[t] hence non zero constants in k. Since $ch_A(t)$ is monic, the coefficient of the leading term on the left equals det(A)det(B). Since the product $\prod f_i$ on the right is also monic, $\det(A)\det(B) = 1$, hence $ch_A(t) = \prod f_i$. **QED.**

Note the analogy here with the structure of finite abelian groups. If G is an abelian group isomorphic to $(Z/n_1) \times \ldots \times (Z/n_r)$, where n_i divides n_{i+1}, then n_r is the annihilator of G, (it generates the principal annihilator ideal), and the cardinality of the group G is $\prod n_i$. In both cases it is hard to compute the precise annihilator, but we can compute a multiple of it more easily, i.e. in one case the order of the abelian group, and in the other the characteristic polynomial of the matrix. In both cases the computable element has the same prime factors as the annihilator.

Next we recall the root - factor theorem, and apply it to prove the theorem above, that the characteristic matrix of A gives a presentation for the k[t] module (k^n, A). We also get another proof of Cayley Hamilton.

Polynomials with non commutative coefficients: If R is any ring, not necessarily commutative, define the polynomial ring R[t] as usual, but where powers of t commute with all coefficients in R, although the coefficients may not commute among themselves.

Hence $f(t) = \sum a_i t^i = \sum t^i a_i$, but if we set t = c, where c is in R, it makes a difference whether we set t = c in the first or the second of these expressions. We call $f_r(c) = \sum a_i c^i$ the right value of f at c, and $f_l(c) = \sum c^i a_i$, the left value of f at c.

Remainder theorem: If f(t) is a polynomial in R[t], then we can write $f(t) = (t-c)q(t) + f_l(c) = p(t)(t-c) + f_r(c)$, i.e. we can divide f(t) by (t-c) from the left, with remainder the left value of f at c, and similarly from the right. The quotients and remainders are unique if we require the remainder belong to R.
proof: We do it for left evaluations and left division. This is the binomial theorem, i.e. replace t in f(t), by (t-c)+c and expand. We get in each term $t^i a_i$, terms in which all but the last have a factor of (t-c), i.e. $t^i a_i = [(t-c)+c]^i a_i = [(t-c)q(t) + c^i] a_i$. Thus $f(t) = \sum t^i a_i = (t-c)Q(t) + \sum c^i a_i$, and we see the remainder is indeed the left evaluation of f at c.

This proves existence. For uniqueness, assume f(t) = (t-c)q(t)+r = (t-c)(p(t)+s, where r,s belong to R. Then (t-c)[q(t)-p(t)] = s-r. Thus the left hand side also belongs to R. But multiplication by (t-c) raises the degree by one, so the left hand side has degree ≥ 1, unless [q(t)-p(t)] = 0. then also r-s = 0. Hence both quotient and remainder are unique. **QED.**

Corollary: If f(t) is any polynomial in R[t], f is left divisible by (t-c) if and only if $f_l(c) = 0$. Similarly for right divisibility.
proof: The expression we gave shows that $f(t) = (t-c)q(t) + f_l(c)$, Hence if $f_l(c) = 0$, then f is left divisible by (t-c). Conversely, if f is left divisaible by (t-c), uniqueness shows the remainder, which is zero, must equal $f_l(c)$, so $f_l(c)$ 0. **QED.**

To apply these results to products of matrices, we prove that matrices with polynomial entries are equivalent to polynomials with matrix coefficients.

Lemma: If k is a field, the non commutative ring $Mat_n(k[t])$ of n by n matrices with entries from k[t], is isomorphic to $Mat_n(k)[t]$, the ring of polynomials with coefficients in the non commutative ring $Mat_n(k)$.
proof: Just as with commutative rings, a ring map R[t]-->S is obtained from a ring map R--->S plus a choice of element in S to send t to, only this time, since t commutes with R in R[t], we must choose as image of t, an element that commutes with the image of R in S. So we map $Mat_n(k)$ into $Mat_n(k[t])$ by viewing scalar matrices as polynomial matrices, and then send t to

the matrix t.I, which is in the center of $Mat_n(k[t])$, i.e. it commutes with everything. It is an exercise to check this ring map is injective and surjective. **QED.**

It follows we we get equivalent results by multiplying two matrices of polynomials as matrices, or as polynomials with matrix entries.

Corollary: Cayley Hamilton. A square matrix A over a commutative ring R, is a root of its characteristic polynomial $ch_A(t)$.
proof: By Cramer's rule, we have $(tI-A).adj(tI-A) = ch_A(t).I$, as products of matrices. Then it holds also as products of polynomials. Setting t = A gives zero on the left, hence also on the right side. I.e. if $ch_A(t) = \sum t^i c_i$, where the c_i belong to R, then $ch_A(t).I = (\sum t^i c_i).I = \sum t^i(c_i.I)$. Thus setting t = A gives $0 = \sum A^i(c_i.I) = \sum A^i(c_i) = \sum c_i A^i = ch_A(A)$. **QED.**

If in the lemma above, we think of the matrix on the left acting individually on each column vector of the matrix on the right, we can also consider matrices of polynomials acting on column vectors of polynomials, as multiplication from the left of polynomials with matrix coefficients, times polynomials with column vector coefficients. I.e. the lemma also holds, with the same proof, for polynomials with coefficients in any ring R with identity, acting from the left on polynomials with coefficients in any (unitary) left module over R.

So let $k^n[t]$ denote polynomials with coefficients which are column vectors from k^n. This is not a ring, in particular the coefficents do not have an element 1, so this object does not contain t. But the coefficients do contain the basic vectors e_i, and we can multiply these by polynomials over k and add up. In particular this object is a k[t] module, and is isomorphic as such to the free k[t] module $(k[t])^n$.

I.e. if E_i are the standard free k[t] basis vectors in $(k[t])^n$, just send E_i to e_i, and $\sum f_i E_i$ to $\sum f_i e_i$ where f_i are polynomials in k[t]. The expression $\sum f_i e_i$ can be re - expanded as a polynomial in t with vector coefficients by expanding each term as $fe_i = (a_0+a_1t+...+t^n)e_i = (a_0e_i + t\,a_1e_i +...+ t^ne_i)$, and then combining coefficients of like powers of t, from various terms, to get coefficient vectors.

Exercise: Show this gives a k[t] module isomorphism $(k[t])^n \to k^n[t]$.

As we have remarked above, the previous lemma, shows multiplication of matrices corresponds to multiplication of polynomials, i.e. the isomorphisms above, give isomorphisms of multiplication diagrams with matrix multiplication $Mat_n(k[t]) \times (k[t])^n \to (k[t])^n$, corresponding to polynomial multiplication $Mat_n(k)[t] \times k^n[t] \to k^n[t]$.

Now we can prove the main presentation theorem.

Theorem: Given any n by n matrix A over a field k, defining a k[t] module structure on k^n, the

k[t] module map $(k[t])^n$--->k^n, sending $\sum f_i(t)E_i$ to $\sum f_i(A)e_i$, is surjective, and its kernel is a free k[t] module, freely generated by the columns of [tI-A], the characteristic matrix of A. I.e. this sequence is exact: 0--->$(k[t])^n$--->$(k[t])^n$--->k^n--->0, as k[t] - modules, where the left map is multiplication by [tI-A].

proof: We know the last map is surjective.

Recall the right map takes $\sum f_i(t)E_i$ to $\sum f_i(A)e_i$, which is exactly the result of viewing $\sum f_i(t)E_i$ as a polynomial $\sum f_i(t)e_i$ with coefficient vectors in k^n, and then setting t = A. So if we view these as maps of polynomials $k^n[t]$--->$k^n[t]$--->k^n--->0, the right map $k^n[t]$--->k^n, is left evaluation of a polynomial f(t) with vector coefficients, at t = A. By the factor theorem above, this is zero if and only if f(t) is left divisible by (t-A), i.e. if and only if f(t) is in the image of the left hand map $k^n[t]$--->$k^n[t]$.

Since multiplication by a monic polynomial never sends a non zero polynomial to zero, the left map is injective. Hence the sequence 0--->$(k[t])^n$--->$(k[t])^n$--->k^n--->0 is exact, and (tI-A) is indeed a presentation matrix for the module (k^n,A). **QED.**

The following amazing theorem, generalizes the fact a surjective endomorphism of a finite dimensional vector space is also injective.

Theorem: If R is any commutative ring and X a finitely generated R module, any surjective R module map f:X--->X is an isomorphism.

proof: This follows from the proof of Cayley Hamilton. If $x_1,...,x_n$ are generators and if we write $f(x_j) = \sum_i a_{ij} x_i$, then as in a previous proof, the matrix A represents f for the generators $\{x_i\}$ even if not independent, and look at the matrix $M = \begin{bmatrix} A-a_{11} & -a_{12} & ... & -a_{1n} \\ -a_{21} & A-a_{22} & ... & -a_{2n} \\ .. & .. & ... & .. \\ -a_{n1} & -a_{n2} & ... & A-a_{nn} \end{bmatrix}$.

Again the transpose

$M^t = \begin{bmatrix} A-a_{11} & -a_{21} & ... & -a_{n1} \\ -a_{12} & A-a_{22} & ... & -a_{n2} \\ .. & .. & ... & .. \\ -a_{1n} & -a_{2n} & ... & A-a_{nn} \end{bmatrix}$, annihilates the column of vectors $\begin{bmatrix} e_1 \\ e_2 \\ ... \\ e_n \end{bmatrix}$

Again the determinant of tI-A is a polynomial P(t) over R annihilating the matrix A and hence the map f. As a small refinement: note if the image f(X) of the map f lies in the submodule IX, for some ideal I of R, then we can choose the entries a_{ij} to belong to I, and looking at the determinant formula for P shows the coefficient of t^i in P(t) belongs to the power I^{n-i} of the ideal I, where n = degree of P(t).

Now apply the principle just proved, not to f, but to the map Id:X--->X where X is viewed not as an R module, but as an R[t] module where t acts via t = f. Then the image of Id is

all of X, which equals (t)X, the product of X by the ideal (t) in R[t]. Hence we have a polynomial satisifed by Id as follows: $Id^n + c_1 f.Id^{n-1} + ...+c_{n-1}f^{n-1}.Id + c_n f^n = 0$, where each $c_i f^i$ belongs to the ideal (f) in R[f]. But we can solve this for Id, getting $Id = -[c_1 f.Id^{n-1} + ...+c_{n-1}f^{n-1}.Id + c_n f^n] = f[-c_1.Id^{n-1} - .....-c_{n-1}f^{n-2}.Id - c_n f^{n-1}]$. The polynomial expression on the right is a right inverse for f, and since all its terms are polynomials in f, it commutes with f, hence is also a left inverse. **QED.**

We have not said how to find bases for k^n which put a given matrix A into rational canonical form. Although in theory one could presumably do it by keeping track of the diagonalization steps, as they do in DF, this seems unappealing. I.e. given A there exist invertible Q such that $(Q^{-1}AQ)$ is in canonical form, but it seems tedious to find such Q in practice.

We will undertake this job only in the simplest possible case, i.e. for nilpotent matrices. We will find it already quite tedious enough there. It is worthwhile however, as in that case it leads to finding so called "Jordan" matrices, a slight variation on the rational canonical form.

Nilpotent matrices, and Jordan canonical forms

Since the companion matrix of a polynomial contains the coefficients of the minimal polynomial of the matrix, it will be as simple as possible when those coefficients are all zero. That happens if and only if the minimal poynomial is t^r for some r, i.e. for operators T such that $T^r = 0$, for some r. We call these operators nilpotent. The Jordan form is a trick to produce a matrix for a general operator in terms of rational canonical matrices of nilpotent operators.

It only works when the characteristic polynomial of T factors completely into linear factors over the field k, but every field has an extension where this holds, as we will learn soon. In fact universal such field extensions of k exist where every polynomial over k factors completely, e.g. the complex numbers do this for matrices over Q or R.

Rational canonical form of a cyclic nilpotent matrix

Let T:V--->V be a linear operator whose associated k[t] module structure is isomorphic to the cyclic module $k[t]/(t^n)$, with annihilator t^n. Then T is nilpotent of index n = dim(V). In the standard k basis $\{[1], [t], [t^2], ...,[t^{n-1}]\}$ for the vector space $k[t]/(t^n)$, the matrix for t has the following rational canonical form, say when n = 5:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

This nilpotent matrix M, with $M^5 = [0]$, corresponds to the module $k[t]/(t^5)$, with annihilator t^5. Equivalently, there is some basis for V, in which T has

this same matrix. This is about as simple as a rational canonical matrix can get. We want to extend the range of this observation.

A cyclic Jordan block

Suppose T:V--->V defines a module structure isomorphic to the cyclic module $k[t]/((t-c)^n)$, almost as simple as before, with annihilator $(t-c)^n$. In spite of the similarity of these two cases, the rational canonical matrix is now quite terrible, being the following matrix for n = 5:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & -c^5 \\ 1 & 0 & 0 & 0 & 5c^4 \\ 0 & 1 & 0 & 0 & -10c^3 \\ 0 & 0 & 1 & 0 & 10c^2 \\ 0 & 0 & 0 & 1 & -5c \end{bmatrix}$$

. This will never do. But the solution is almost obvious.

Namely, we should have looked at the matrix for (T-c) instead of the matrix for T. I.e. if T satisfies the polynomial $(t-c)^n$, then (T-c) satisfies the polynomial t^n, i.e. (T-c) is nilpotent even though T is not. So we should have taken the rational canonical matrix for T-c, instead of for T. This means we get the module structure $k[X]/(X^n)$ for V, where multiplication by X is action by T-c, hence the standard basis $\{[1], [X], [X^2],...,[X^{n-1}]\}$ for $k[X]/(X^n)$, corresponds to the basis $\{[1], [(t-c)], [(t-c)^2], ... , [(t-c)^{n-1}]\}$ for $k[t]/((t-c)^n)$.

The latter is a cyclic basis for (T-c), and in that basis the (rational canonical) matrix for (T-c) is the standard nilpotent matrix above for n=5. But we want a matrix for T, not (T-c). This however is trivial, since T = cI+(T-c). Now cI is so simple it has a diagonal matrix in any basis at all, so if we use the rational canonical basis for (T-c) just chosen, in that basis, T has the following Jordan matrix (where n=5):

$$\begin{bmatrix} c & 0 & 0 & 0 & 0 \\ 1 & c & 0 & 0 & 0 \\ 0 & 1 & c & 0 & 0 \\ 0 & 0 & 1 & c & 0 \\ 0 & 0 & 0 & 1 & c \end{bmatrix}$$

This is the <u>sum</u> of the rational canonical matrix for (T-c),

plus the rational canonical matrix for cI. A more general Jordan matrix is composed of blocks like this.

A nilpotent matrix with more than one block

If T:V--->V defines a module structure on V isomorphic to a product of cyclic modules of form $(k[t]/(t^{r}1))$ x ....x $(k[t]/(t^{r}m))$, then T is nilpotent of index $r = r_m$, i.e. t^r annihilates the module, and is the minimal polynomial for T. Then the rational canonical matrix for T consists of exactly m blocks of nilpotent cyclic matrices, of sizes $r_1,...,r_m$. E.g. the following illustrates the case $(k[t]/(t^2))$ x $(k[t]/(t^3))$:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$ Note it has lower rank than the nilpotent matrix M above.

A matrix of Jordan blocks all with the same eigenvalue

The analog of the previous nilpotent matrix is an operator T with minimal polynomial $(t-c)^r$, where $r < n = \dim(V)$, and as a module V is a product $[k[t]/((t-c)^{r_1})] \times \ldots \times [k[t]/((t-c)^{r_m})]$, with $r_1 \leq r_2 \leq \ldots \leq r_m = r$. The Jordan matrix for this T is obtained from that for the nilpotent version, by putting c's everywhere on the diagonal. If $V \approx k[t]/(t-c)^2 \times k[t]/(t-c)^3$, we have the following Jordan matrix for T:

$$\begin{bmatrix} c & 0 & 0 & 0 & 0 \\ 1 & c & 0 & 0 & 0 \\ 0 & 0 & c & 0 & 0 \\ 0 & 0 & 1 & c & 0 \\ 0 & 0 & 0 & 1 & c \end{bmatrix}$$. Note this matrix has full rank if $c \neq 0$.

In general, if the minimal polynomial for T is $\prod (t-c_i)^{r_i}$, there will be some blocks with c_1 on the diagonal, the largest of which are of size r_1, some blocks with c_2 on the diagonal the largest being of size r_2, etc... The only thing to be determined is how many blocks exist of each size, for each eigenvalue c.

Exercise: a) If the minimal polynomial is $\prod(t-c_i)$, and the characteristic poloynomial is $\prod(t-c_i)^{s_i}$, the Jordan matrix of is diagonal, with s_i diagonal entries equal to c_i for each i.

b) Conversely, if the Jordan matrix is diagonal, and the characteristic polynomial is $\prod(t-c)^{s_i}$, then the minimal polynomial is $\prod(t-c_i)$.

c) If there is only one Jordan block for each eigenvalue, then the characteristic and minimal polynomials are equal.

Our discussion shows the following:

Theorem: If T:V--->V has characteristic polynomial $ch_T(t) = (t-c)^n$, i.e. if there is only one characteristic root, then (T-c) is nilpotent, and the matrix of T in some basis consists of one or more Jordan blocks, all with c on the diagonal. The determinant of T is c^n, and the trace is nc.

Existence of Jordan form

The general theorem is the following:

Theorem: Let T:V--->V is a linear endomorphism of a finite dimensional vector space V over k,

whose characteristic polynomial factors completely into linear factors, i.e. $ch_T(t) = \prod(t-c_i)^{r_i}$, i=1,...,m; with all roots $c_1,...,c_m$ in k. Then V is a product of m subspaces $V = V_1 x....x V_m$, such that the restriction T_i of T to each V_i, has form $T_i = c_iI + N_i$ where N_i is nilpotent. Consequently, there is a basis for each V_i, and hence for V, in which the matrix of T is composed of Jordan blocks.

Proof: If m = 1, i.e. there is only one eigenvalue, we have already seen this is true, i.e. no decomposition of V is needed. So assume there are at least two distinct linear factors of $ch_T(t) = \prod(t-c_i)^{r_i}$, i=1...m.

Define V_i = {those w in V such that $(t-c_i)^{r_i}(w) = 0$} = the generalized eigenspace of T for the eigenvalue c_i. Define the natural map $\prod V_i$--->V taking $(w_1,...,w_m)$ to $\sum w_i$. We claim this is an isomorphism. This is an easy Euclidean algorithm argument as follows.

Let $Q_i = \prod(t-c_j)^{r_j}$, $i \neq j$. Since $m \geq 2$, there are at least two Q_j, and they have no common prime factor in k[t]. Hence the ideal they generate in the pid k[t] is the unit ideal, so there exist polynomials P_i such that $\sum P_iQ_i = 1$. Now suppose taking $(w_1,...,w_m)$ is in the kernel of the map above. Then $\sum w_i = 0$, so w_1 is a linear combination of the w_j's with j > 1. But w_1 is annihilated by $(t-c_1)$, hence by all Q_j with j > 1; and all w_j with j > 1 are annihilated by Q_1. Hence $(\sum P_iQ_i)(w_1) = 0$. But $\sum P_iQ_i = 1$, so $(\sum P_iQ_i)(w_1) = w_1$. Similarly we see all $w_i = 0$, and our map is injective.

For surjectivity, let w be any element of V and apply the equation $\sum P_iQ_i = 1$, to w, (using commutativity), getting $\sum(Q_iP_i)(w) = \sum(P_iQ_i)(w) = w$. But $Q_i(u)$ is annihilated by $(t-c_i)^{r_i}$, so any vector in the image of Q_i belongs to V_i. Thus the equation $\sum(Q_iP_i)(w) = w$, shows every vector w is a sum of vectors $(w_1,....,w_m)$ from the product $\prod V_i$. Hence our map is surjective.

QED.

Now the existence of Jordan form is proved, and the uniqueness also follows from the uniqueness for rational form. If we subtract off the diagonal terms from the Jordan form for T, we get a nilpotent matrix, so it follows tool that every T whose minimal polynomial has only linear factors, can be written as a sum T = S+N where, S is diagonalizable, N is nilpotent, and in fact SN = NS. Such S and N are also unique.

At last we bring up the task we have been avoiding, actually computing a basis which puts T into Jordan form. We will start with the easiest case, nilpotent matrices and see why it is already not so easy.

Look at the module $(V,T) \approx k[t]/(t^n)$ and consider the standard cyclic basis for the standard model, namely $\{[1],[t],....,[t^{n-1}]\}$. We started with the generator [1], and just applied t to it n-1 times. So to get a cyclic basis for V, all we need is a T cyclic generator corresponding to [1]. But how to find one? I.e. we need a vector w in V such that $T^nw = 0$, but $T^{n-1}w \neq 0$. So we need to find a basis for the space $\ker T^{n-1}$, and then choose a vector w not belonging to it.

It is not always obvious how to choose a vector not belonging to the span of a given set,

although statistically speaking, over R say, any vector chosen at random has probability 1 of working. An algorithm that will always work is to let $v_1,.....,v_s$ be given, and let $e_1,...,en$ be a basis for the whole space. Then reduce the ordered spanning set $\{v_1,...,v_s,e_1,...,e_n\}$ to a basis starting from the left and discarding any vector which depends on those to its left. Then any e_j remaining afterwards will not be in the span of the v's. This can be done with matrices by Gaussian elimination.

So to find a cyclic generator for V in case the minimal polynomial of T is t^n and n = dimV, we find a basis $v_1,...,v_{n-1}$ for $\ker T^{n-1}$, and then use some such procedure to find a vector w not in that kernel. Then the rational canonical matrix is obtained from the cyclic basis $\{w,Tw,T^2w,...,T^{n-1}w\}$. Similarly, if the minimal polynomial is $(t-c)^n$, we find a basis of $\ker(T-c)^{n-1}$, and then find a vector not in that kernel.

Of course we usually have a more complicated situation in practice, since we may only know the characteristic polynomial and it may not equal the minimal polynomial. So we have more than one cyclic block, hence we are looking for more than one cyclic generator, and the cyclic blocks have different sizes.

We proceed as follows, by example: suppose ker(T-c) has dimension d_1, then there are exactly d_1 Jordan blocks with c on the diagonal. If $\ker(T-c)^2$ has dimension d_1+d_2, then $d_2 \leq d_1$, and there are exactly d_2 blocks of size larger than one, hence exactly d_1-d_2 blocks of size one. If $\ker(T-c)^3$ has dimension $d_1+d_2+d_3$, then $d_3 \leq d_2$, there are exactly d_3 blocks of size greater than 2, hence exactly d_2-d_3 blocks of size two.....Once the dimension of $\ker(T-c)^r$ equals that of $\ker(T-c)^{r+1}$, then there are no blocks larger than r, and we are done with determining the sizes of the blocks for the eigenvalue c. We still need to show how to compute a Jordan basis for this generalized eigenspace.

Here are some actual prelim problems of this nature:

"Put the following matrix in Jordan form":

$$T = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -3 \\ -1 & -1 & 2 \end{bmatrix}.$$ We need to know the characteristic polynomial. With a 3by3 it should be feasible to compute it directly by the determinant. Ok, I got $ch(t) = (t+1)(t-2)^2$. Since (t+1) occurs with multiplicity one, there is exactly one block, and it has size one. Solving for vectors in the kernel of (T+1), gives [1,-1,0], by Gausian elimination. This is a keeper.

$$T-2 = \begin{bmatrix} -1 & 2 & 3 \\ 1 & -2 & -3 \\ -1 & -1 & 0 \end{bmatrix}.$$ We need a vector in this kernel.

It has rank 2, so again get only one, e.g. [1,-1,1]. Hence this is not a cyclic generator for this subspace, since we want a vector annihilated by $(T-2)^2$ but not by (T-2). So we square (T-2), getting:

$(T-2)^2 = \begin{bmatrix} 0 & -9 & -9 \\ 0 & 9 & 9 \\ 0 & 0 & 0 \end{bmatrix}$, so a basis for the kernel, is {[1,0,0], [0,1,-1]}. We want one which does not lie in ker(T-2). Applying T-2 to the first, it is not zero, so we take it. Now we have our cyclic generalized eigenvector, so our cyclic basis for that space is [1,0,0], and (T-2)([1,0,0]) = [-1,1,-1]. Thus our Jordan basis seems to be {[1,-1,0]; [1,0,0], [-1,1,-1]}. The matrix Q with these as columns should conjugate our matrix T into Jordan form, if all is well.

I.e. we should have $Q^{-1}TQ = J = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$, or another ordering of blocks.

Multiplying gives $TQ = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -3 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ -1 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 & -2 \\ 1 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix}$.

Now $QJ = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & -2 \\ 1 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix}$. Since TQ = QJ, thus $J = Q^{-1}TQ$.

Now let $A = \begin{bmatrix} 0 & 0 & -1 & 2 \\ 1 & 1 & 1 & -3 \\ 1 & 0 & 2 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and try it again. We want the characteristic roots, but this time lets diagonalize the characteristic matrix.

$(tI-A) = \begin{bmatrix} t & 0 & 1 & -2 \\ -1 & t-1 & -1 & 3 \\ -1 & 0 & t-2 & 3 \\ 0 & 0 & 0 & t-1 \end{bmatrix}$. Diagonalizing it, yields the following:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (t-1) & 0 \\ 0 & 0 & 0 & (t-1)^3 \end{bmatrix}$$. Thus we have the characteristic polynomial = $(t-1)^4$,

the minimal polynomial = $(t-1)^3$, and we know the Jordan form has one cyclic block of size 1, and one cyclic block of size 3. So we need to find two independent vectors u,v in ker(T-1). Then we need to find a vector w in $\ker(T-1)^3$ that is not in $\ker(T-1)^2$. Then we check to see which of u,v, is niot sdependent on $(T-1)^2w$, say u is not. Then our Jordan basis will consist of {u, w, (T-1)w, $(T-1)^2w$}. I.e. {u, $(T-1)^2w$} is then a basis for ker(T-1), while {u, (T-1)w, $(T-1)^2w$} is a basis for $\ker(T-1)^2$. The increase in dimension by one, from 2 to 3, tells us there is exactly one block of size 2 or more. Then {u, w, (T-1)w, $(T-1)^2w$} is our full Jordan basis for V = $\ker(T-1)^3$. We need to solve the equations of course.

Since it already midnight, this is an exercise.

8000 spectral theorems.

A spectral theorem is a theorem guaranteeing certain special operators or matrices are diagonalizable, without having to actually diagonalize them, or even calculate any characteristic polynomials. Roughly, "symmetric" operators are always diagonalizable.

A hermitian product on a complex vector space is a bi - additive pairing VxV-->C with values in the complex numbers C, such that v.w = conj(w.v) ["conjugate symmetric"], and (av).w = a(v.w) and v.(bw) = conj(b)(v.w). So the pairing is one and a half times complex linear, or "sesquilinear". It follows that for all v in V, v.v = (conj(v.v) is real. We assume v.v > 0 for all non zero v, i.e. the pairing is positive definite.

A hermitian pairing thus defines a length by |v|^2 = v.v. We call the space V equipped with the given hermitian pairing, a hermitian space.

An endomorphism A:V-->V is "hermitian" for a given pairing if (Av).w = v.(Aw) for all v,w in V. A pair of vectors is orthogonal for the given pairing if v.w = 0. Then we have:

Spectral theorem (complex hermitian case)

Theorem: If A:V-->V is a hermitian endomorphism on a finite dimensional hermitian space V, then V has an orthogonal basis of eigenvectors for A, hence also an orthonormal such basis.

Indeed this is more than we need to assume. The more general statement is this: consider the map V-->Hom(V,C) defined by sending v to ( ).v. This is well defined since the pairing is complex linear in the left variable. Moreover, since v.v = 0 implies v = 0, the map is injective, and conjugate linear, hence bijective. Thus every linear map f in Hom(V,C) determines a unique vector v such that f(w) = w.v for all w. In particular, given any endomorphism A, and any vector v, the linear map f(w) = (Aw).v has form w.y for some unique y, which we call A*v. Then the function sending v to A*v, is complex linear. So A* is an endomorphism called the (hermitian) adjoint of A.

Then we claim that the conclusion of the theorem holds if and only if AA* = A*A, and call such A normal. I.e. the operators that commute with their hermitian adjoints are exactly the ones that admit orthonormal eigenbases. The easy direction is that if there is an orthonormal basis in which A has diagonal matrix, then in that same basis the matrix of A* is also diagonal, hence these matrices commute, so also A and A* commute as operators in any basis.

Spectral theorem (normal case)
Theorem: If A:V-->V is a normal endomorphism on a finite dimensional hermitian space V, then V has an orthogonal basis of eigenvectors for A, hence also an orthonormal such basis.
Proof:

Lemma: If A is normal, then A and A* have the same kernel.
proof: Assume Ax = 0. To show A*x = 0 it suffices to show its length is zero. But (A*x).(A*x) = x.(AA*x) = x.(A*Ax) = (Ax).(Ax) = 0. So A*x = 0. **QED.**

Cor: If A is normal, with eigenvalue c, then A maps the ortho - complement of the eigenspace ker(A-c) into itself.
proof: If A is normal, then so is (A-c), as one checks directly. Then ker(A-c) = ker(A*-c*) where c* = complex conjugate of c. Hence if x is in ker(A-c), then also (A*-c*)x = 0. Thus if y is orthogonal to x, i.e. if y.x = 0, then (A-c)y.x = y.(A*-c*)x = 0, so (A-c)y is also orthogonal to x. But then 0 = (A-c)y.x = Ay.x - cy.x = Ay.x - 0, so Ay.x = 0 too. I.e. Ay is also orthogonal to x. **QED.**

Now we construct an orthogonal eigenbasis for V by induction. I.e. the characteristic polynomial for A has a complex root and hence there is a corresponding eigenvector x1, which we may scale down to length one. Now if we restrict A to the ortho complement of <x1>, the restriction of A is again normal and maps that smaller subspace into itself. Thus by induction there is an orthonormal eigenbasis of this smaller subspace. Combined with x1, we have an orthonormal eigenbasis for V. **QED.**

Examples: An operator A is called hermitian if A = A*, unitary if A^(-1) = A*, and a real operator is called symmetric if A = A* for some real inner product. Hermitian and unitary operators are examples of normal operators, hence they have orthonormal eigenbases.

Remark: If A is hermitian, the characteristic polynomial of A has all real roots, i.e. ch(A) always splits over the reals.
proof: If Av = cv, then c(v.v) = (cv).v = (Av).v = v.(Av) = v.(cv) = c*(v.v), and since $v \neq 0$, c = c*, so c is real.

A real symmetric operator is in a sense an example of a hermitian operator on the "complexification" of the real space, and the theorem holds again. But we give an independent proof in the real case, from my linear algebra notes. The only difference is in the first step, where we cannot use the characteristic polynomial to rpoduce the first eigenvector, since we do not know it has a real root.

The transpose of a matrix, symmetric matrices.

Defn: An n by n matrix A is **symmetric** if the entry in the i^{th} column and j^{th} row equals the entry in the j^{th} column and i^{th} row, for every i and j.
The matrix A* obtained from A by interchanging its rows and columns is called the **transpose of A**. Thus A is symmetric if and only if A = A*.

Ex: i) If the operations are defined, (A+B)* = A* + B*, and (AB)* = B*A*.
ii) If A is any m by n matrix, and v, w are any vectors in R^n, R^m respectively, then v.(A*w) = (Av).w.
iii) If A = A*, then Av.w = v.Aw, for all v,w.

Spectral theorem (real symmetric case)

Thm: If A = A*, R^n has a basis of mutually orthogonal eigenvectors of A.

Pf: The real valued function f(x) = Ax.x has a maximum on the unit sphere in R^n, at some point y where the gradient df of f is "zero", i.e. df(y) is perpendicular to the tangent space of the sphere at y. The tangent space at y is the subspace of vectors in R^n perpendicular to y, and df(y) = 2Ay. Hence Ay is perpendicular to the tangent space at y, i.e. Ay = 0 or Ay is parallel to y, so Ay = cy for some c, and y is an eigenvector for A.

Now restrict A to the subspace V of vectors orthogonal to y. If v.y = 0, then Av.y = v.Ay = v.cy = c(v.y) = 0. Hence A preserves V. A still has the property Av.x = v.Ax on V, so the restriction of A to V has an eigenvector in V. (Although V has no natural representation as R^{n-1}, the argument for producing an eigenvector depended only the symmetry property Av.x = v.Ax.) Repeating, A has an eigenbasis. **QED.**

Dual spaces and pairings

One can discuss pairings in a more abstract setting, without making any choices, simply by considering vectors in a space V over a field k, together with "covectors" or dual vectors, i.e. elements of the function space V* = Hom(V,k) = {k linear functions V→k}. Here there is a natural pairing, not between elements of the same space. But between pairs of elements of V and V*. I.e. by the very definition of elements of V* there is a natural"evaluation" pairing VxV*→k, taking (v,f) to f(v). The function space V* is called the "dual space of V".

To add to the confusion, we can consider the double dual V** = (V*)*, and the natural map V→V**, taking v to v** = "evaluation at v", which of course is a linear map v**:V*→k, taking f to v**(f) = f(v). If $v \neq 0$, it follows from choosing a basis of V containing v, that there is some f with $v^{**}(f) \neq 0$, namely f = projection of a vector w on its v- coefficient in that basis. Thus the map V→V** is always injective, and hence also surjective in case V has finite k dimension. Thus in fin ite dimensions, V and V** are essentially the same, although this is not true for V and V*. I.e. although we will see that V and V* have the same dimension when those are finite, there is no natural way to identify them, and they deserve to be distinguished.

It is "obvious" that V and V* are different since there is a natural pairing between them, so if V and V* were naturally the same there would always be a natural pairing of V with itself. On the

other hand, once we specify a pairing on V, then V and V* do become related, or even identified for some purposes, by means of that pairing.

I.e. suppose we are given a pairing VxV→k which is bilinear, i.e. for every u,v,w in V, and a in k, we have <u+v,w> = <u,w> + <v,w>, <u,v+w> = <u,v> + <u,w>, and <av,w> = a <v,w> = <v,aw>. Then the map V→V* taking v to < . ,v>, is linear. If the pairing is non degenerate, in the sense that <u,v> = 0 for all u only when v = 0, the map V→V* is injective, hence also bijective if V has finite k dimension, which allows us to identify V and V* for some purposes.

For example, if V is an n dimensional real vector space, and we have a usual "inner product" on V, i.e. a symmetric, bilinear, "positive definite" pairing [i.e. v.v > 0 for all v ≠ 0], this yields an isomorphism V→V*, taking v to ( ).v.

[If V has infinite k dimension, then I believe V* has still larger infinite dimension, and in particular they are never isomorphic in the infinite dimensional case. Thus also V and V** are not isomorphic when V has infinite dimension, if I am correct. (I have not written the proof.)]

Conversely if we are given a linear map V→V* taking v to the linear function ( ).v, we can consider this a pairing on V, by defining <w,v> = (w).v. This pairing will be bilinear, but not necessary symmetric or non degenerate. So it seems that bilinear pairings VxV→k are equivalent to linear maps V→V*, and non degenerate pairings are equivalent to linear injections V→V*. We will discuss below how to recognize those maps V→V* that correspond to symmetric, or skew symmetric, pairings.

Abstract orthogonal complements and transposes
Given a vector space V over a field k, and a subspace W of V, define the orthogonal complement Wperp in V* to be the subspace of those linear functions in V* which are identically zero when restricted to W. As usual, these are naturally identified with precisely all linear functions in (V/W)*, i.e. a linear function V/W→k defines by composition one V→V/W→k, and conversely a linear function V→k which annihilates W induces one V/W→k. So Wperp is a subspace of V* which is naturally isomorphic to (V/W)*.

If V has finite dimension, then dim(V) = dim(W) + dim(V/W) = dim(W) + dim(V/W)*
= dim(W) + dim(Wperp), which is the usual formula for the dimensions of a subspace and its orthogonal complement in Euclidean space.

There is also an abstract analog of the transpose of a map T:V→W, namely T* is the map T*:W*→V* taking a linear function f:W→k in W*, to the composition (foT):V→W→k in V*. Confusing isn't it? Anyway, T* just means "precede by T". That's often what upper star means, in algebraic topology, functional analysis, and elsewhere.

Do this again and, from T:V→W, you get a map T** = (T*)*:V**→W**. Now remember the natural map V→V*? That means whenever you have a map T:V→W, that now you also have maps V→W→W**, and V→V**→W**, and we claim you can check these compositions are equal. In category language, this says there is a "natural transformation" between the identity and the double dual operations (such operations are called functors there).

It now makes sense to speak of symmetric or skew symmetric operators in an abstract sense, at least for the special case of maps $T:V \to V^*$. I.e. then $T^*:V^{**} \to V^*$, so this provides a natural map $V \to V^{**} \to V^*$ to compare with $T:V \to V^*$. If they are equal we say the map T was symmetric, or if they are negatives of each other, call T skew symmetric.

I guess it will follow that bilinear pairings $V \times V \to k$ that are symmetric (or skew symmetric) in the usual sense, yield maps $V \to V^*$ which are symmetric (or skew symmetric) in the abstract sense, and vice versa. You might try checking that, to see if you have some facility with the definitions.

Dual bases and matrices

To see how the abstract transpose relates to the matrix transpose, we have to introduce bases. So let V be finite dimensional with basis $v_1,\ldots,v_n$. Then we can define linear functions $f_1,\ldots,f_n$ on V, by setting $f_1 = 1$ on v_1 and $= 0$ on the other basis elments, then setting $f_2 = 1$ on v_2, and $= 0$ on the other basis elements, and so on. These n functions are independent an span V*, and are called the basis for V* dual to the original basis for V. Note that we must know the whole basis for V to define even one of the f's, so the individual elements of the basis for V* are not dual just to the corresponding element of the basis for V, but the whole basis is dual to the whole other basis. Thus although sometimes we write the basis dual to $v_1,\ldots,v_n$ as $v_{1*},\ldots,v_{n*}$, this is a bit misleading, since for example v_{1*} depends on $v_1,\ldots,v_n$ and not just v_1.

Anyway a basis for V, also defines a map $V \to V^*$ sending each v_i to v_{i*}. Thus this map defines a pairing on V. We claim the basis $v_1,\ldots,v_n$ is orthonormal for this pairing. Vice versa, if we have a pairing on V and consequently a map $V \to V^*$, we can ask whether that map sends a given basis for V to its dual basis, and the answer should be that it does if and only if the basis is orthonormal for the given pairing?

Now if $T:V \to W$ is a linear map of finite dimensional spaces with matrix A in some bases for V and W, then $T^*:W^* \to V^*$ also has a matrix in terms of the corresponding dual bases, which we claim is the usual transpose of A. To see that, just recall how you find the elements of a matrix in terms of a basis: namely the i,j entry is the ith coeficient of the image of the jth basis vector. So let the basis of V be denoted by v's, that of W be w's, and the dual bases be f's and g's. Thus the i,j entry of the matrix for T* is the ith coefficient of $T^*(g_j)$, i.e. $T^*(g_j)(v_i) = g_j(T(v_i))$, which is the jth coefficient of the image of the ith basis vector under T. So the i,j entry of A* is the j,i entry of A. See if you can convince yourself of that.

Suppose we have a map $T:V \to V^*$ which is symmetric in the abstract sense, that is: $T = T^*:V^{**}=V \to V^*$. Should the spectral theorem hold? What would it say? That there is a basis for V such that T becomes diagonal in terms of that basis and its dual basis? Presumably not, or there would be such an abstract version in every book. The first step would be to show that if $T \neq 0$, then for some v, T(v) is not zero on v, so that T(v) could be a multiple of v* for some basis of V containing v. Can you find a counterexample?