845 course notes part 2a

§8) Primary decomposition and Jordan canonical form

Whenever the minimal polynomial μ of a k-linear endomorphism T factors completely into linear factors, as always happens when k is an algebraically closed field such as $\mathbb{C}$, there is a slight variation of the rational canonical form which gives a simpler matrix for T, the so-called Jordan canonical form. Rational canonical form is simplest when the operator is nilpotent, i.e. when some power of it is zero, since then the coefficients of the minimal polynomial appearing along the right side are all zero. To exploit this for other operators, we first decompose the space according to the distinct prime factors of μ, and in each factor space corresponding to a factor $(X-\lambda)^r$ of μ, choose the rational canonical basis corresponding to the nilpotent operator $(T-\lambda)$, rather than T. Since we must add λ's along the diagonal to pass from the matrix of $T-\lambda$ to the matrix of T, the union of these bases gives a basis for the whole space such that the corresponding matrix for T is composed of blocks of form:

$$\begin{bmatrix} \lambda & 0 & 0 & . & 0 \\ 1 & \lambda & 0 & . & 0 \\ 0 & 1 & \lambda & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & 1 & \lambda \end{bmatrix}$$: λ's on the diagonal, 1's below it, 0's elsewhere.

To calculate the number of blocks for each λ, and the size of each block, we have two options, either find and then diagonalize, a "presentation matrix" $[\varphi]$ for the $k[X]$ module (M,T), or else carry out the kind of dimension calculations used in the proof of uniqueness of a cyclic decomposition. We will show how to do such calculations in the next section. In this section we will describe the Jordan form and prove it exists. It is traditional to use λ instead of X for the variable, to speak of (M,T) as a $k[\lambda]$-module, and the minimal polynomial $\mu(\lambda)$ as a polynomial in λ. In general however, it seems easier to keep the variables separate from the constants if we use X for the variable in the polynomial $\mu(X)$, and λ or λ_i for the roots

Let $T:M\to M$ be an endomorphism of a finite dimensional k-vector space, with minimal polynomial $\mu = \Pi_{j=1,\dots,s}(X-\lambda_i)^{r_i}$, i.e. assume all

roots of μ are in the field k. Thus the prime factors of μ have form $p_i = (X-\lambda_i)$, for $i = 1,\ldots,s$. Our first step is a preliminary decomposition of M into the "torsion" subspaces determined by powers of each prime factor. This is the analog of the decomposition of a finite abelian group into a product of its Sylow subgroups, but here it is called the "primary decomposition" of (M,T). This step does not need the hypothesis that $\mu(X)$ factors into linear factors over k, and can be done with any finite dimensional pair (M,T).

Primary decomposition Lemma: Given a (finite dimensional) pair (M,T), let $\mu(X) = \Pi p_i^{r_i}$ be the factorization of the minimal polynomial of T into powers of distinct irreducible factors over k. Define the "primary" subspaces of M as follows: for each i, let $M_i = \{v \text{ in } M: (p_i(T))^{r_i}(v)=0\}$ = {all vectors killed by the ith factor $(p_i(T))^{r_i}$ of μ, }. Then $M_i \subset M$ is a T-invariant subspace, and $M \cong \Pi_i M_i$, isomorphic as k-vector spaces (and as k[X] - modules).
proof: The fact that each M_i is T-invariant follows from the fact that T commutes with any polynomial in T such as each $p_i(T)$. I.e. if v is in M_i then $(p_i(T))^{r_i}(v)=0$; then $(p_i(T))^{r_i}(Tv) = T((p_i(T))^{r_i}(v)) = T(0) = 0$, so Tv is in M_i too.
The decomposition follows from a generalization of the argument for the Chinese remainder theorem as follows: since each $M_i \subset M$ is a subspace, we can map the product $\psi: \Pi_i M_i \to M$ by addition, i.e. if $a = (a_1,\ldots,a_s)$, define $\psi(a) = \Sigma a_i$. Since the M_i are T-invariant and $T(\Sigma a_i) = \Sigma T(a_i)$, this is a k[T] module map. To see ψ injective, assume $\psi(a) = \Sigma a_i = 0$, so $a_1 = -a_2-\ldots-a_s$ lies in $M_2+\ldots+M_s$. Then a_1 is annihilated by both $p_1^{r_1}$, and by $(p_2\cdot\ldots\cdot p_s)^t$ for some $t > 0$, (since the right side is). However, since $p_1^{r_1}$ and $(p_2\cdot\ldots\cdot p_s)^t$ have no common prime factors, their gcd is 1, and we can write $1 = g\cdot p_1^{r_1} + h\cdot(p_2\cdot\ldots\cdot p_s)^t$. Then 1 annihilates a_1, so $a_1 = 0$. A similar argument shows all $a_i = 0$, hence ψ is injective

To show ψ surjective, let $\mu = \Pi_{i=1\ldots s} p_i^{r_i}$ again be the factorization of the minimal polynomial, and note that the various products $q_1 = (p_2^{r_2}\cdot\ldots\cdot p_s^{r_s})$, $q_2 = (p_1^{r_1}\cdot p_3^{r_3}\cdot\ldots\cdot p_s^{r_s})$, ..., $q_s = (p_1^{r_1}\cdot\ldots p_{s-1}^{r_{s-1}})$, have no common prime factors, where q_i is formed by omitting the power $p_i^{r_i}$ from the minimal polynomial. Hence together they

generate the unit ideal, and we can write $1 = h_1q_1+\ldots+h_sq_s$, for some polynomials h_j. Then for each element b in M, we have $b = 1\cdot b = \Sigma h_jq_jb = \Sigma b_j$, where $b_j = h_jq_jb$. If we note $p_i^{r_i}\cdot b = h_im(X)b = 0$, we see that b_i belongs to M_i. This proves surjectivity of ψ. QED.

After decomposing M according to the primary decomposition $M \cong \Pi_{i=1\ldots s} M_i$, we can then decompose each of the k[X] modules M_i by the standard decomposition. If we write T_i for the restriction of T to M_i, the resulting matrix for T is made up of s blocks, where the ith block is the rational canonical matrix of T_i. Thus each $[T_i]$ is itself composed of blocks, in which each block is the companion matrix of some power of p_i

The two decompositions, primary and standard, can also be done in the other order. If we have already done the standard decomposition of (M,T), $M \cong \Pi_\alpha\ k[X]/(f_\alpha)$, we can get the primary decomposition in two more steps: first, take the primary decomposition of each cyclic factor; then for each prime factor p of the minimal polynomial μ of T, the p-primary subspace M_p of M (already in its standard decomposition), is the product of the p-primary subspaces of the various cyclic factors $k[X]/(f_\alpha)$ of M.

The primary decomposition of a cyclic factor is a corollary of the lemma above, but is perhaps worth stating in a more general form, which is just as easy to prove:
Sublemma: If R is any ring and $\{I_j\}$ a finite collection of "comaximal" ideals, in the sense that $I_j+I_k = R$ for all $j\neq k$, then the natural map $\varphi: R \to \Pi_j\ (R/I_j)$ induces an isomorphism $R/(\cap_j I_j) \to \Pi_j\ (R/I_j)$.
proof: To show the natural map $\varphi: R \to \Pi_j\ (R/I_j)$ is surjective, it suffices to show the image contains the standard generators $(0,\ldots 0,[1],0\ldots..,0)$ For each j we must produce an element x in R such that $x \equiv 1$ mod I_j, and x is in I_k for all $k\neq j$. Given j, for each $k\neq j$ choose x_k in I_k and y_k in I_j so that $x_k+y_k = 1$. Then set $x = \Pi_k x_k = \Pi_{k\neq j}(1-y_k) = 1 + y$, where y is in I_j. Thus $\varphi(x) = (0,\ldots 0,[1],0\ldots..,0)$, and φ is surjective. Since $\varphi(z) = 0$ iff z is in every I_j, $\ker(\varphi) = \cap_j I_j$. Thus by the first isomorphism theorem, $R/(\cap_j I_j) \cong \Pi_j\ (R/I_j)$. QED.

Corollary: If R is a p.i.d., μ a non unit, $\mu = \Pi_i\ p_i^{r_i}$, for non associate primes p_i, then $R/(\mu) \cong \Pi_i\ R/(p_i^{r_i})$.

proof: The ideals (p^r) and (q^s) are comaximal when p,q, are non associate primes in R. I.e. since $\gcd(p^r,q^s) = 1$, they generate the ideal $(p^r,q^s) = (1) = R$. Note also, if $\mu = \Pi p_i^{r_i}$, then $\cap_i (p_i^{r_i}) = (\mu)$. Thus $R/(\mu) \cong R/(\cap_i (p_i^{r_i})) \cong \Pi\ R/(p_i^{r_i})$. QED.

Example of primary decomposition

By way of illustration, if we use the same example as in the previous section where the standard decomposition of (M,T) was $(k[X]/(X-2)) \times (k[X]/((X-2)(X-3)^2) \times k[X]/((X-2)^2(X-3)^2)$, the invariant factors are $f_1 = X-2$, $f_2 = (X-2)(X-3)^2$, $f_3 = (X-2)^2(X-3)^2$, and the prime factors are, $p = X-2$, and $q = X-3$. Then each of the three cyclic factors can be decomposed into its primary components as follows:

The first cyclic factor $(k[X]/(X-2))$ is already $(X-2)$-primary, i.e. its annihilator is a power of the prime $X-2$. The second cyclic factor decomposes as $(k[X]/((X-2)(X-3)^2) \cong (k[X]/(X-2)) \times (k[X]/(X-3)^2)$, and the third cyclic factor decomposes as $k[X]/((X-2)^2(X-3)^2) \cong (k[X]/(X-2)^2) \times (k[X]/(X-3)^2)$. Then the $(X-2)$-primary subspace of M is the product of those for each of the three factors, i.e. $M(X-2) \cong (k[X]/(X-2)) \times (k[X]/(X-2)) \times (k[X]/(X-2)^2)$.

The $(X-3)$-primary component of M is likewise the product of those of each cyclic factor. Of course only two of the cyclic factors have them, so we get $M(X-3) \cong (k[X]/(X-3)^2) \times (k[X]/(X-3)^2)$. Then $M \cong M(X-2) \times M(X-3)$, and we already have the standard decomposition of each of those submodules. The matrix of T for this decomposition is then composed of five blocks, namely the companion matrices for the polynomials, $(X-2)$, $(X-2)$, $(X-2)^2$, $(X-3)^2$, and $(X-3)^2$. This looks as follows:

$$\begin{bmatrix} 2 & & & & & & \\ & 2 & & & & & \\ & & 0 & -4 & & & \\ & & 1 & 4 & & & \\ & & & & 0 & -9 & \\ & & & & 1 & 6 & \\ & & & & & & 0 & -9 \\ & & & & & & 1 & 6 \end{bmatrix} = [T].$$

The first three blocks along the diagonal correspond to the matrix of T on the subspace M(X-2), and the next two blocks give the matrix of T on the subspace M(X-3). There is really no big advantage to this matrix over the other version of the rational canonical form, but if we make a slight change in our point of view, then we get something very nice, as we discuss next.

Jordan canonical form

Suppose we look again at the map T on the subspace M(X-2). On this space the map (T-2) has an even simpler minimal polynomial than does T, namely the minimal polynomial of (T-2) is just X^2. More generally, whereas T has invariant factors (X-2), (X-2), $(X-2)^2$ on this subspace, (T-2) has the even simpler invariant factors X, X, X^2. I.e. these polynomials have only zeroes as non leading coefficients! Hence the companion matrices are as follows:

$[0] = [CX]$, $[0] = [CX]$, $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = [CX^2]$, and we have the following rational canonical matrix for (T-2), on M(X-2).

$$\begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 0 & 0 \\ & & 1 & 0 \end{bmatrix} = [T\text{-}2], \text{ on } M(X\text{-}2).$$

Now since T = 2I + (T-2), and since in any basis 2I has a diagonal matrix with 2's on the diagonal, the rational canonical basis for (T-2) [on M(X-2)], gives this matrix for T:

$$\begin{bmatrix} 2 & & & \\ & 2 & & \\ & & 2 & 0 \\ & & 1 & 2 \end{bmatrix} = [T], \text{ on } M(X-2), \text{ where the blanks are zeroes.}$$

Similarly, on $M(X-3)$, $(T-3)$ has minimal polynomial X^2, and invariant factors X^2, X^2, so we get the following matrix for $(T-3)$, on $M(X-3)$:

$$\begin{bmatrix} 0 & 0 & & \\ 1 & 0 & & \\ & & 0 & 0 \\ & & 1 & 0 \end{bmatrix} = [T-3], \text{ on } M(X-3).$$

Hence, as matrix for T on $M(X-3)$ we get:

$$\begin{bmatrix} 3 & 0 & & \\ 1 & 3 & & \\ & & 3 & 0 \\ & & 1 & 3 \end{bmatrix} = [T], \text{ on } M(X-3).$$

Combining these gives the full Jordan matrix for T.

$$\begin{bmatrix} 2 & & & & & \\ & 2 & & & & \\ & & 2 & 0 & & \\ & & 1 & 2 & & \\ & & & & 3 & 0 \\ & & & & 1 & 3 \end{bmatrix} = [T], \text{ on all of } M.$$

This last matrix is the Jordan canonical form for T, and it is unique, except for the possible reordering of the roots 2,3 of the minimal polynomial. (Recall that $\mathbb{C}$ is not an ordered field, so there is no natural way to order these roots, at least when they include non real complex numbers, although a non natural order may be agreed

upon.) Thus, while the rational canonical form of [T] is given by choosing the following k-bases for the five cyclic primary subspaces: {1}, {1}, {1,X}, {1,X}, {1,X}, the Jordan canonical form arises from choosing instead the following k-bases for those same subspaces: {1}, {1}, {1, (X-2)}, {1, (X-3)}, {1, (X-3)}.

In general, after decomposing (M,T) according to primary subspaces, and assuming the minimal polynomial factors completely over k into linear factors $\mu = \Pi_{i=1...s} (X-\lambda_i)^{t_i}$, and supposing that on the subspace $M(X-\lambda)$, T has invariant factors $(X-\lambda)^{r_1}, (X-\lambda)^{r_2},...,(X-\lambda)^{r_n}$, where $r_1 \leq ... \leq r_n$, then in each of the corresponding cyclic subspaces $k[X]/(X-\lambda)^r$, we choose as basis $\{1,(X-\lambda),...,(X-\lambda)^{r-1}\}$. This will give us a matrix $[J_\lambda]$ for [T] on $M(X-\lambda)$ composed of n blocks, one block for each r_i. Each block has λ along the diagonal and 1's below the diagonal, and the block corresponding to $(X-\lambda)^r$ is an $r \times r$ matrix.

To recap, an "elementary Jordan block" is a square matrix of the following form, $r \times r$ for some r, with some root λ of μ along the diagonal, ones just below the diagonal, zeroes elsewhere:

$$\begin{bmatrix} \lambda & & & \\ 1 & \lambda & & \\ & 1 & \lambda & \\ & & \ddots & \ddots \\ & & & 1 \quad \lambda \end{bmatrix} = J_{r,\lambda}$$

Then, if the $(X-\lambda)$ subspace of (M,T) has n cyclic components, $M_\lambda \cong \Pi_{i=1...n} k[X]/(X-\lambda)^{r_i}$, the matrix $[J_\lambda]$ of T on M_λ will contain exactly n such elementary Jordan λ-blocks:

$$\begin{bmatrix} [J_{r_1,\lambda}] & & & \\ & [J_{r_2,\lambda}] & & \\ & & \ddots & \\ & & & [J_{r_n,\lambda}] \end{bmatrix} = [J_\lambda].$$

Finally if $\mu = \Pi_{i=1...s} (X-\lambda_i)^{t_i}$, the full matrix for T will be made up of s such block matrices J_λ, one for each root λ of the minimal polynomial $\mu(X)$ for T.

$$\begin{bmatrix} [J_{\lambda_1}] & & \\ & [J_{\lambda_2}] & \\ & & [J_{\lambda_s}] \end{bmatrix} = [T], \text{ on all of M.}$$

It follows from the theory of standard decomposition that the Jordan canonical form of T exists and is unique, up to ordering of the roots λ of $\mu(X)$, assuming only the minimal polynomial μ splits over k. We can derive a few useful corollaries about diagonalizable endomorphisms from the Jordan form as follows.

Definition: An endomorphism is "diagonalizable" iff in some basis, it has a matrix all of whose off-diagonal entries are zero.

Note that a diagonal matrix is in Jordan form, so an endomorphism is diagonalizable iff it has a Jordan form and that form is diagonal (i.e. "all λ's and no 1's").

Corollary: T is diagonalizable iff all the standard invariant factors of (M,T) are square free, iff the minimal polynomial μ_T is square free. In particular, if M has dimension = d over k, and if μ has d distinct roots, $\lambda_1, \ldots, \lambda_d$, then $\mu = \Pi_{i=1...d} (X-\lambda_i)$, and T is diagonalizable.
proof: Exercise in the definition of the Jordan form. QED.

Exercise #142) (i) Write down all possible 5×5 Jordan matrices having minimal polynomial $(X-5)^2$.
(ii) Write down all possible 5×5 Jordan matrices having minimal polynomial $(X-1)(X-3)(X+6)$.
(iii) Write down all possible 6×6 Jordan matrices having minimal polynomial $(X-1)^2(X-2)^2$.
(iv) Write down the Jordan form of T if the pair (M,T) has invariant factors $(X-1)(X-2)$, $(X-1)^3(X-2)$, $(X-1)^3(X-2)^2(X-5)^3$.

§9) The canonical presentation of (M,T).

We want to show how to calculate the Jordan canonical form of an endomorphism $T:M \to M$, of a finite dimensional vector space M, starting from any matrix for T. We already know how to do it starting from the invariant factors of the pair (M,T), so one way to proceed is to find a "presentation matrix" for (M,T), i.e. a matrix for the map representing M as a quotient of two free k[X] modules. Then we can diagonalize that matrix by row and column operations to find the invariant factors, and thus the Jordan form. Of course, as a practical matter, we also have to be able to split the invariant factors into linear factors over k, which can be challenging.

Definition: Let (M,T) be given, where M is a finite dimensional k vector space, and $T:M \to M$ an endomorphism. If F_1, F_2 are finite rank, free k[X] modules, and $\varphi: F_1 \to F_2$ is a k[X] module map, such that for some k[X] map $F_2 \to M$ the sequence $F_1 \to F_2 \to M \to 0$ is exact as k[X] modules, then a matrix for φ is called a "presentation matrix" for (M,T).

Terminology: For any map $\varphi: A \to B$, the quotient $B/Im(\varphi)$ is called the "cokernel of φ", or $coker(\varphi)$.

Remark: Note that $\varphi: F_1 \to F_2 \to M$ is a presentation for M, iff $F_2 \to M$ induces $F_2/Im(\varphi) \cong M$ Thus a presentation of M is an isomorphism of M with a cokernel of a map of free modules.

Given any matrix for T, there is always a canonical associated presentation matrix for (M,T) as follows:

Theorem: Let $M \cong k^m$ by some choice of k-basis for M, and let $A = [a_{ij}]$ be the associated matrix for T. Since elements of k are also elements of k[X], the matrix A in $Mat_{m \times m}(k)$ also belongs to $Mat_{m \times m}(k[X])$. Then the matrix $[X \cdot I - A]$ in $Mat_{m \times m}(k[X])$ is a presentation matrix for (M,T).

Proof: Let the standard k[X]-basis for $(k[X])^m$ be denoted $u_1, \ldots, u_m$. Thus $u_1 = (1,0,\ldots,0)$, $u_2 = (0,1,0,\ldots,0)$, ... etc. We denote the standard k-basis for k^m by $e_1, \ldots, e_m$. Then $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0)$, ... etc. These bases $\{u_j\}$ and $\{e_j\}$ thus look the same but we distinguish

them since they belong to different spaces and play different roles for us. Note the inclusion $k^m \subset (k[X])^m$ identifies e_i with u_i.
Of course k^m is also a $k[X]$ module, where for v in k^m, $X \cdot v = Av$. Then $\{e_j\}$ is a $k[X]$ generating set, but not a $k[X]$-basis, for k^m. Thus the map $\sigma: (k[X])^m \to k^m$ taking $u_j \mapsto e_j$ is a $k[X]$ surjection. Note that $\sigma(X \cdot u_j) = X \cdot \sigma(u_j) = X \cdot e_j = Ae_j = j$th column of $A = (a_{1j}, a_{2j}, \ldots, a_{mj})$. If $f = (f_1, \ldots, f_m) = \Sigma_i f_i u_i$, then $\sigma(f) = \Sigma_i f_i(A)e_i$.
Define the $k[X]$-map $\varphi: (k[X])^m \to (k[X])^m$ by the matrix $[\varphi] = [X \cdot I - A]$. Then $\varphi(u_j) = j$th column of $[\varphi] = (-a_{1j}, -a_{2j}, \ldots, X - a_{jj}, \ldots, -a_{mj}) = X \cdot u_j - \Sigma_i a_{ij} u_i$.
Now apply σ to this last expression, $\sigma(\varphi(u_j)) = \sigma(X \cdot u_j - \Sigma_i a_{ij} u_i) = X \cdot e_j - \Sigma_i a_{ij} e_i = Ae_j - (\Sigma_i a_{ij} e_i) = Ae_j - Ae_j = 0$. Hence $\mathrm{Im}(\varphi) \subset \ker(\sigma)$.

The following assertion implies the theorem:
Claim: The sequence $0 \to (k[X])^m \to (k[X])^m \to k^m \to 0$ defined by the maps φ and σ is exact.
proof: We have already shown σ is surjective and $\sigma \circ \varphi = 0$. Since k^m has rank zero as $k[X]$ module, by additivity of rank, $\mathrm{Im}(\varphi)$ has rank m, $\ker(\varphi)$ has rank zero, so $\ker(\varphi) = \{0\}$ and φ is injective. Hence it suffices to show $\mathrm{Im}(\varphi) \subset \ker(\sigma)$. Assume f is in $\ker(\sigma)$, i.e. $\sigma(f) = \Sigma_i f_i(A)e_i = 0$, in k^m. We must show f is in $\mathrm{Im}(\varphi)$.
Lemma: (i) $\Sigma_i f_i(A)u_i = 0$ also in $(k[X])^m$.
(ii) For any polynomial $h(X)$, $h(X) \cdot I = h(X \cdot I)$.
(iii) For any polynomial h, $h(X \cdot I) - h(A) = (X \cdot I - A) \cdot g$, for some g in $\mathrm{Mat}_{m \times m}(k[X])$.
proof of lemma: **(i)** By hypothesis, $0 = \sigma(f) = \Sigma_i f_i(A)e_i$ in k^m, and the inclusion $k^m \subset (k[X])^m$ identifies the vector $\Sigma_i f_i(A)e_i$ with the vector $\Sigma_i f_i(A)u_i$. QED.
(ii) Whenever R is a ring, a, b are in R, and S, T are R-homomorphisms we always have $(aS)(bT) = (ab)(ST)$, so in particular in our case where $R = k[X]$, we have $(X \cdot 1)(X \cdot I) = X^2 \cdot I$, $(X \cdot I)^n = X^n \cdot I^n = X^n \cdot I$, ..., so taking k-linear combinations of these gives $h(X) \cdot I = h(X \cdot I)$, for any h. QED.
(iii) Although $\mathrm{Mat}_{m \times m}(k[X])$ is not a commutative ring, $X \cdot I$ is in the center of this ring. Hence for all n, we have the usual factorization $(X \cdot I)^n - A^n = (X \cdot I - A)(X^{n-1} \cdot I + X^{n-2} \cdot I \cdot A + \ldots + A^{n-1})$.

Since $h(X\cdot I)-h(A)$ is a k-linear combination of differences such as $(X\cdot I)^n - A^n$, it follows that $(X\cdot I - A)$ is a factor of each term and hence of the difference $h(X\cdot I)-h(A)$. **QED. Lemma**

We can now show that f is in $Im(\varphi)$. I.e. then $f = \Sigma_i f_i(X)\cdot u_i = \Sigma_i f_i(X)\cdot I\cdot u_i = \Sigma_i f_i(X\cdot I)\cdot u_i - \Sigma_i f_i(A)\cdot u_i = \Sigma_i (f_i(X\cdot I) - f_i(A))\cdot u_i = \Sigma_i (X\cdot I - A)g_i\cdot u_i = (X\cdot I - A)(\Sigma_i g_i\cdot u_i) = \varphi(\Sigma_i g_i\cdot u_i)$, is in $Im(\varphi)$.
QED for Claim and Theorem.

Terminology: We call $[X\cdot I - A]$ the characteristic matrix for A.

Examples: We now give examples of applying this theorem to find the Jordan canonical form of a matrix. Suppose the matrix whose Jordan form is wanted is the following:

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 6 & 1 \\ -16 & -16 & -2 \end{bmatrix}. \text{ Then } X\cdot I - A = \begin{bmatrix} X-2 & 0 & 0 \\ -4 & X-6 & -1 \\ 16 & 16 & X+2 \end{bmatrix}. \text{ Interchanging}$$

rows and columns to bring the smallest $\neq 0$ entry to upper left gives:

$$\begin{bmatrix} -1 & X-6 & -4 \\ 0 & 0 & X-2 \\ X+2 & 16 & 16 \end{bmatrix}. \text{ Adding } (X+2)\cdot(\text{top row}) \text{ to the bottom row,}$$

gives the following matrix:

$$\begin{bmatrix} -1 & X-6 & -4 \\ 0 & 0 & X-2 \\ 0 & (X-2)^2 & -4(X-2) \end{bmatrix}. \text{ Clearing out the top row, and then}$$

exchanging the resulting two right hand columns gives:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & X-2 & 0 \\ 0 & -4(X-2) & (X-2)^2 \end{bmatrix} \text{ Adding 4 times middle row to bottom row:}$$

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & X-2 & 0 \\ 0 & 0 & (X-2)^2 \end{bmatrix}$, which is in standard diagonal form.

Thus the invariant factors are (X-2), $(X-2)^2$. There is only one irreducible factor, X-2, which occurs twice, with exponents 1 and 2. Thus there are two elementary Jordan blocks for the characteristic root X=2, one of size 1×1, and one of size 2×2. Thus the Jordan form is the following matrix:

$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$ = Jordan form of A.

Let's try another example:

$B = \begin{bmatrix} 0 & 1 & -2 & 1 \\ -2 & 1 & -6 & 3 \\ 2 & -3 & 0 & 1 \\ 2 & -3 & -2 & 3 \end{bmatrix}$. Then the characteristic matrix is:

$X \cdot I - B = \begin{bmatrix} X & -1 & 2 & -1 \\ 2 & X-1 & 6 & -3 \\ -2 & 3 & X & -1 \\ -2 & 3 & 2 & X-3 \end{bmatrix}$. Bringing as small as possible an

entry to the upper left position gives the following matrix:

$\begin{bmatrix} 1 & 2 & -1 & X \\ 3 & 6 & X-1 & 2 \\ 1 & X & 3 & -2 \\ 3-X & 2 & 3 & -2 \end{bmatrix}$. Using entry (1,1) to clear the first column:

$\begin{bmatrix} 1 & 2 & -1 & X \\ 0 & 0 & X+2 & 2-3X \\ 0 & X-2 & 4 & -2-X \\ 0 & 2X-4 & 6-X & X^2-3X-2 \end{bmatrix}$ Clear 1st row, then bring 4 to (2,2),

and concentrate on lower right 3×3 submatrix:

$$\begin{bmatrix} 4 & X-2 & -X+2 \\ X+2 & 0 & 2-3X \\ 6-X & 2X-4 & X^2-3X-2 \end{bmatrix}$$. Clear 1st column, then 1st row:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4-X^2 & X^2-8X+12 \\ 0 & X^2-4 & 3X^2-8X+4 \end{bmatrix}$$. Now concentrate on the lower right 2×2

submatrix, and reduce degree of new entry (1,1) by adding 2nd column to 1st:

$$\begin{bmatrix} -8X+16 & X^2-8X+12 \\ 4X^2-8X & 3X^2-8X+4 \end{bmatrix} = \begin{bmatrix} 8(2-X) & (X-2)(X-6) \\ 4X(X-2) & 3X^2-8X+4 \end{bmatrix}$$. Now entry (1,1)

is the gcd of the 1st row and 1st column, so clear them both:

$$\begin{bmatrix} 8(2-X) & X^2-8X+12 \\ 0 & X^3-2X^2-4X-8 \end{bmatrix} \approx \begin{bmatrix} X-2 & 0 \\ 0 & (X-2)^2(X+2) \end{bmatrix}$$. The standard diagonal

form is thus $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & X-2 & 0 \\ 0 & 0 & 0 & (X-2)^2(X+2) \end{bmatrix}$. Now we know the two

invariant factors to be {(X-2), $(X-2)^2(X+2)$}. Then we have two characteristic roots, $\lambda = -2, 2$ There is only one invariant factor divisible by a power of (X+2), with exponent 1, hence only one Jordan block corresponding to $\lambda = -2$, of size 1×1.
There are two invariant factors containing powers of (X-2), with exponents 1 and 2, hence there are two Jordan blocks for $\lambda = 2$, of sizes 1×1 and 2×2 We have this Jordan matrix:

$$\begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$ ≈ Jordan form for B.

Now let's try a still larger one:

$$C = \begin{bmatrix} 5 & -1 & -3 & 2 & -5 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & -2 \\ 0 & -1 & 0 & 3 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{bmatrix}$$. The characteristic matrix is:

$$X\cdot I - C = \begin{bmatrix} X-5 & 1 & 3 & -2 & 5 \\ 0 & X-2 & 0 & 0 & 0 \\ -1 & 0 & X-1 & -1 & 2 \\ 0 & 1 & 0 & X-3 & -1 \\ -1 & 1 & 1 & -1 & X-1 \end{bmatrix}$$. Exchange 1st, 5th rows,

and clear 1st column, then 1st row:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & X-2 & 0 & 0 & 0 \\ 0 & -1 & X-2 & 0 & 3-X \\ 0 & 1 & 0 & X-3 & -1 \\ 0 & X-4 & X-2 & 3-X & X^2-6X+10 \end{bmatrix}$$. Focus on lower right 4×4 and

exchange new rows 1 and 3:

$$\begin{bmatrix} 1 & 0 & X-3 & -1 \\ -1 & X-2 & 0 & 3-X \\ X-2 & 0 & 0 & 0 \\ X-4 & X-2 & 3-X & X^2-5X+6 \end{bmatrix}$$. Clear 1st column, then 1st row:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & X-2 & X-3 & 2-X \\ 0 & 0 & -X^2+5X-6 & X-2 \\ 0 & X-2 & -X^2+6X-9 & X^2-5X+6 \end{bmatrix}$$. In lower right 3×3, subtract

new 2nd column from 1st:

$$\begin{bmatrix} 1 & X-3 & 2-X \\ X^2-5X+6 & -X^2+5X-6 & X-2 \\ X^2-5X+7 & -X^2+6X-9 & X^2-5X+6 \end{bmatrix}$$ Subtract 3rd row from 2nd,

then clear 1st column, 1st row, then look at lower right 2×2:

$$\begin{bmatrix} 0 & -X^2+5X-6 \\ -X^3+7X^2-16X+12 & X^3-8X^2+23X-22 \end{bmatrix}$$ Switch columns and add

X·(1st row) to 2nd row, then negate first row:

$$\begin{bmatrix} X^2-5X+6 & 0 \\ -3X^2+17X-22 & -X^3+7X^2-16X+12 \end{bmatrix}$$ Add 3· (1st row) to 2nd row,

then factor first row:

$$\begin{bmatrix} (X-2)(X-3) & 0 \\ 2(X-2) & -X^3+7X^2-16X+12 \end{bmatrix}$$ Add (3-X)·(second row) to

2·(first row), and then switch rows:

$$\begin{bmatrix} 2(X-2) & -X^3+7X^2-16X+12 \\ 0 & X^4-10X^3+37X^2-60X+36 \end{bmatrix}$$ Clear out first row, and factor:

$$\begin{bmatrix} 2(X-2) & 0 \\ 0 & (X-2)^2(X-3)^2 \end{bmatrix}$$ Now we have the diagonal matrix:

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & (X-2) \\ & & & (X-2)^2(X-3)^2 \end{bmatrix}$$ This gives the invariant

factors { (X-2), $(X-2)^2(X-3)^2$}. Thus there are two characteristic roots, $\lambda = 2, 3$ Two powers of (X-2) occur with exponents 1 and 2, so there are two Jordan blocks for $\lambda = 2$, of sizes 1×1 and 2×2. The only power of (X-3) that occurs has exponent 2 so there is one

Jordan block for $\lambda = 3$, and it has size 2×2. Thus the Jordan matrix is the following:

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix} = \text{Jordan matrix for C.}$$

In the next section we introduce the characteristic polynomial of an endomorphism T, describe it using determinants, and show how to calculate a basis which puts [T] in Jordan form.

Exercise #143) Find the Jordan forms of these matrices:

(i) $A = \begin{bmatrix} 0 & -1 & 2 \\ 3 & -4 & 6 \\ 2 & -2 & 3 \end{bmatrix}$. (ii) $B = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$. (iii) $C = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ -4 & 1 & -3 & 2 & 1 \\ -2 & -1 & 0 & 1 & 1 \\ -3 & -1 & -3 & 4 & 1 \\ -8 & -2 & -7 & 5 & 4 \end{bmatrix}$.

845 course notes part 2b

§10) Characteristic polynomials, eigenvectors, and Jordan bases

In dealing with finite abelian groups, Cauchy's theorem told us that the order (i.e. cardinality) of a group is a multiple of the annihilator of the group. This was useful information since the order of a group is often more easily calculated than the annihilator. In the present setting of a k[X] module (M,T) the minimal polynomial of T is the analog of the annihilator of the group. There is also an analog for (M,T) of the order of the group, namely the characteristic polynomial χ, a polynomial which may be more readily calculated than the minimal polynomial μ, and which turns out to be a multiple of μ. χ also contains other useful information about T, such as $\dim_k(M)$, det(T), and trace(T), and the dimensions of all the primary components of M

Definition: Given a square matrix A, define the "characteristic polynomial" of A to be $\chi_A = \det(X \cdot I - A)$.

Remarks: If A is an $m \times m$ matrix, χ_A is a monic polynomial of degree m. Similar matrices have the same characteristic polynomial.

proof: [second statement only]: For any two $m \times m$ matrices C,D, $\det(CD) = \det(C)\det(D)$, and $C^{-1}(X \cdot I)C = X \cdot I$. Thus if $B = C^{-1}AC$, then $(X \cdot I - B) = (X \cdot I - C^{-1}AC) = C^{-1}(X \cdot I)C - C^{-1}AC = C^{-1}(X \cdot I - A)C$. Thus $\chi_B = \det(X \cdot I - B) = \det(I)\det(X \cdot I - B) = \det(C^{-1}C)\det(X \cdot I - B) = \det(C^{-1})\det(C)\det(X \cdot I - B) = \det(C^{-1})\det(X \cdot I - B)\det(C) = \det(C^{-1}(X \cdot I - B)C) = \det(X \cdot I - C^{-1}BC) = \det(X \cdot I - A) = \chi_A$ **QED.**

Corollary: If $T: M \to M$ is an endomorphism of a finite dimensional k-vector space, we may define the characteristic polynomial of T uniquely by $\chi_T = \chi_A$ where A is any matrix for T.

Corollary: If $f_1, \ldots f_s$ are the invariant factors of $T: M \to M$, then $\chi_T = \prod_j f_j$; i.e. the characteristic polynomial of T is the product of the invarient factors of T.

proof: If A is any matrix for T, then the invariant factors are the diagonal entries on the matrix resulting from diagonalizing the

presentation matrix [X·I-A], over the ring k[X]. Hence there are invertible k[X] matrices σ,τ such that $[X\cdot I-A] = \tau\cdot D\cdot\sigma$, where D is a diagonal $m\times m$ matrix whose diagonal entries are possibly some 1's at the beginning, and then the invariant factors $f_1,\ldots,f_s$.

$$D = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & f_1 & \\ & & & & \ddots \\ & & & & & f_s \end{bmatrix}$$

Then $\det(X\cdot I-A) = \det(\sigma)\det(D)\det(\tau)$, where $\det(\sigma),\det(\tau)$ are invertible elements of k[X], i.e. non zero elements of k. Thus on the left we have a monic polynomial of degree m, and on the right we have a constant multiple of a monic degree m polynomial. Hence the constant multiple is 1, and $\chi_T = \det([X\cdot I-A]) = \det(D) = \Pi_i f_i$, as claimed. **QED.**

Corollary(Cayley - Hamilton): T satisfies its characteristic polynomial, i.e. $\chi_T(T) = 0$.
proof: We know $\mu(T) = 0$, where the minimal polynomial μ is the invariant factor of T of highest degree. Since $\mu|\chi$, $\chi(T) = 0$. **QED.**

Remark: Since the characteristic polynomial may be more easily computable than the minimal polynomial, it is useful to have a diagonalizability criterion in terms of it. We do not have a necessary criterion but we do have a sufficient one:

Corollary: Given $T:M\to M$, where $\dim_k(M) = m$, if χ_T has m distinct roots in k, then the Jordan form is diagonal; in particular, T is diagonalizable.
proof: In the $m\times m$ Jordan form there are m distinct blocks each for a different root λ, hence each block is 1×1, and the matrix is diagonal. QED.

Exercise #144) (i) If A is a 3×3 matrix with characteristic polynomial $\chi = (X-4)^3$, give all possible Jordan forms for A, each

with its minimal polynomial.
(ii) If $\chi = \prod_\lambda (X-\lambda)^{m_\lambda}$ is the characteristic polynomial of $T:M \to M$, prove every root of χ is also a root of the minimal polynomial μ, and if $M_\lambda = \{v \text{ in } M: \text{ for some } r > 0, (T-\lambda)^r(v) = 0\}$ is the primary subspace of M corresponding to the root λ, that $\dim(M_\lambda) = m_\lambda$.
(iii) Use determinants to compute χ for these matrices:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}, C = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}, D = \begin{bmatrix} 1 & -2 & -1 & 0 \\ 1 & 0 & -3 & 0 \\ -1 & -2 & 1 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}.$$

Exercise # (45) If A,B are $m \times m$ matrices, over $k = \bar{k}$, define trace(A) = tr(A) = $\Sigma_i a_{ii}$.
(i) Prove $\det(A) = (-1)^m \chi(0) = (-1)^m \cdot$(constant term of χ).
(ii) Prove tr(AB) = tr(BA), deduce $\text{trace}(B^{-1}AB) = \text{trace}(A)$.
(iii) Prove: tr(A) = Σ (roots of χ) = -(coeff. of X^{m-1} in χ).
(iv) If $M_\lambda = \{v \text{ in } k^m : (A-\lambda)^r v = 0 \text{ for some } r > 0\}$, prove $\dim(M_\lambda) = d$, iff $(X-\lambda)^d \mid \chi$, but $(X-\lambda)^{d+1} \nmid \chi$.

Remark: Since the characteristic polynomial of an endomorphism T is an invariant of T, so are all its coefficients, which must equal the elementary symmetric functions of the characteristic roots. The previous exercise shows that the trace and determinant of an endomorphism T are (up to sign) the first and last of these coefficients. We could call all the coefficients of χ "characteristic invariants" of T, but the others don't seem to get used much.

Next we discuss how to find, not just the Jordan matrix for a given A, but the Jordan basis corresponding to this matrix. Equivalently we show how to find an invertible matrix Q such that $Q^{-1}AQ = J$, the Jordan matrix of A. We will focus on finding bases for the subspaces of form $\ker(A-\lambda)^r$ where λ is a root of the characteristic polynomial of A. The first of these is particularly important and has a special name, the "λ-eigenspace of A".

Definition: An "eigenvalue" of A is an element λ of the field k such that $A-\lambda$ has a non trivial kernel. If λ is an eigenvalue, $\ker(A-\lambda)$ is

the corresponding eigenspace. An "eigenvector" of A is a non zero vector v such that $(A-\lambda)v = 0$, for some λ in k. The eigenvectors of A for λ are the non zero elements of $\ker(A-\lambda)$.

Remark: The element λ is an eigenvalue of A iff λ is a root of χ_A, i.e. iff λ is a characteristic root of A.
proof: If $(A-\lambda)v = 0$, where $v \neq 0$, then the matrix $A-\lambda$ is not invertible, hence $\det(A-\lambda) = 0$. I.e. λ is a root of $\det(A-X) = \chi$. Conversely, if λ is a root of χ, then $\det(A-\lambda) = 0$, so $(A-\lambda)$ is not invertible, hence $A-\lambda$ cannot be injective, so $\ker(A-\lambda) \neq \{0\}$. **QED.**

Now suppose $A = \begin{bmatrix} 8 & -4 \\ 9 & -4 \end{bmatrix}$ has Jordan matrix $\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} = J$. How do we find a "Jordan" basis of k^2 corresponding to this matrix?
Note that J-2 has matrix $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, and thus $(J-2)e_1 = e_2$, $(J-2)e_2 = 0$. This says e_2 is an eigenvector for J, with eigenvalue 2. So the basis in which the endomorphism A has matrix J, is of form v_1, v_2, where $(A-2)v_1 = v_2$, $(A-2)v_2 = 0$. Since one thing we know how to do with any matrix C is solve homogeneous equations of type $Cv = 0$, we could begin the search for this basis by finding a vector v_2 such that $(A-2)v_2 = 0$. Then how to find v_1? Well since $(A-2)v_1 = v_2$, and $(A-2)v_2 = 0$, we see $(A-2)^2 v_1 = 0$. Thus we could try to find v_1 by solving $(A-2)^2 v_1 = 0$. Note however that v_1 should not also be a solution of $(A-2)v_1 = 0$. Since there are two independent solutions of $(A-2)^2 v = 0$, but only one independent solution of $(A-2)v = 0$, such a v_1 does exist. So here is what we do: first find w_2 that solves $(A-2)w = 0$. Then w_2 solves $(A-2)^2 w = 0$, but does not span the full solution space of $(A-2)^2 = 0$. Let w_1 be any second solution of $(A-2)^2 = 0$ which is independent of w_2. Thus both of $\{w_1, w_2\}$ solve $(A-2)^2 w = 0$, and only one of them, namely w_2, solves $(A-2)w = 0$. This is not quite the set $\{v_1, v_2\}$ we want, since we also want to have $(A-2)v_1 = v_2$. So put $v_1 = w_1$, but replace w_2 by $v_2 = (A-2)v_1$. Then $\{v_1, v_2\}$ is the basis we want.

So we have to begin this process by finding an eigenvector w_2 in $\ker(A-2)$ which is not the one we ultimately want. We just use w_2

to help us find w_1, in $\ker(A-2)^2 - \ker(A-2)$, which *is* a basis vector we want. I.e. after finding w_1, we take $v_1 = w_1$, and then go back and choose a new $v_2 = (A-2)v_1$. Then $\{v_1,v_2\}$ is our Jordan basis.

Let's actually do the calculation for the matrix A above. Since $A-2 = \begin{bmatrix} 6 & -4 \\ 9 & -6 \end{bmatrix}$, first we solve $\begin{bmatrix} 6 & -4 \\ 9 & -6 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, for $w_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Since $(A-2)^2 = 0$, we just pick any vector independent of w_2 to be w_1, such as $w_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then we take $v_1 = w_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $v_2 = (A-2)v_1 = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$. Then the basis $\{v_1,v_2\} = \{(1,0), (6,9)\}$ is a Jordan basis for A. I.e. since $A(1,0) = (8,9) = 2\cdot(1,0)+1\cdot(6,9)$, the first column of the matrix of A in the basis is $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and since $A(6,9) = (12, 18) = 0\cdot(1,0) + 2\cdot(6,9)$, the second column is $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$. If we now define $Q = \begin{bmatrix} 1 & 6 \\ 0 & 9 \end{bmatrix}$, then $Q^{-1} = (1/9)\begin{bmatrix} 9 & -6 \\ 0 & 1 \end{bmatrix}$, and you can check that $Q^{-1}AQ = J = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$.

When we work with matrices for which we do not know the answer in advance, and especially for large matrices, we will find the numbering of vectors in the basis gets completely mixed up. We have already seen in the previous example that w_2 was the first vector found, because w_2 was a solution of the first equation we wrote down, $(A-2)w = 0$. I.e. w_2 was an eigenvector. This notational problem only gets worse with more complicated matrices. One attempt at a numbering system seems to be to reverse what we have just done, and try to number the basis vectors backwards, i.e. so that the eigenvectors come first. But even this does not work perfectly, since the vectors found first, i.e. the eigenvectors, correspond to the right most columns of each individual Jordan block, so they should be ultimately scattered throughout the basis. Moreover, numbering this way reverses the order of the columns in the Jordan matrix, changing our "lower" jordan form into an "upper" jordan form, but that is really no problem. To illustrate, let's try a 3×3 example, numbering "backwards" this time.

Let's redo the following 3×3 example from the previous section.

$A = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 6 & 1 \\ -16 & -16 & -2 \end{bmatrix}$. We know the Jordan form is $J = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

[Here we may think of A as an endomorphism of $M = \mathbb{Q}^3$.] The eigenvectors of J are e_1 and e_3, so in the first step of the procedure, when we solve $(A-2)w = 0$, we will come up with versions of the first and third basis vectors, and as we saw, these will have to be modified as in the following discussion.

Assume we know the characteristic polynomial is $\chi = (X-2)^3$. Thus the sum of the sizes of the $\lambda = 2$ blocks is 3. I.e. $\dim_{\mathbb{Q}}(M_2) = 3$, where $M_2 = \{v \text{ in } M: \text{ for some } r > 0, (A-2)^r(v) = 0\} = \{v \text{ in } M. (A-2)^3(v) = 0\}$ is the 2-primary component of M. (Of course here $M_2 = M$.) First solve $(A-2)w = 0$, getting two independent solutions α_1, β_1. (The subscript is supposed to be the smallest power of $(A-2)$ which annihilates these vectors.) These are eigenvectors. Since we only found two and not three independent eigenvectors, the minimal polynomial μ is not $(X-2)$. But μ is also not $(X-2)^3$, since then we would have found only one independent eigenvector. Hence μ is $(X-2)^2$, and the jordan form has one 1×1 block and one 2×2 block.

Next solve $(A-2)^2w = 0$, finding one more solution vector α_2 independent of α_1, β_1. Now $\{\alpha_1,\beta_1,\alpha_2\}$ is a basis for $\ker(A-2)^2$, and the first two vectors $\{\alpha_1,\beta_1\}$ are a basis for $\ker(A-2)^1$. Since we have three independent vectors and $\dim(M_2) = 3$, we have a basis of M_2. But we need a "cyclic" basis. I.e. the problem at present is there is no relation between α_1 and α_2, so we cannot take them together as a Jordan basis for the 2×2 block. As usual the one found last, here α_2, is a "keeper", but we must change either α_1 or β_1. So we set $a_2 = \alpha_2$, and define $a_1 = (A-2)a_2$. Then a_1 is a new eigenvector, so it must replace either α_1 or β_1, but we don't know which one yet. Order the vectors like this: $\{a_1,\alpha_1,\beta_1,a_2\}$. This is a generating set for M_2 but not an independent one, since all three vectors with subscript "1" belong to the two dimensional space $\ker(A-2)^1$. So we reduce it to a basis, starting at the left end, working from left to right, eliminating any vectors that depend on

the vectors to their left. (As you probably know, there is a standard procedure for this by row reduction, or "Gaussian elimination".) This will eliminate either α_1 or β_1, and we then relabel the remaining one as b_1. Then we have $\{a_1,b_1,a_2\}$ where $(A-2)a_2 = a_1$.
As a set, this is the Jordan basis, but in the wrong order. If we reorder it, as $\{a_1,a_2,b_1\}$, then the Jordan matrix will be in the "upper Jordan form", i.e. the 1's will be above the diagonal instead of below. That is fine, and we call this basis an upper diagonal Jordan basis. I.e. the Jordan form will look like this:

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$ The larger Jordan block is first instead of last, and

the 1's are above rather than below the diagonal. This is a perfectly good Jordan form. If you want to get our "lower Jordan form" instead, reverse the order of the basis vectors to get $\{a_2, b_1, a_1\}$.

Let's carry all this out in detail:
Since $\chi = (X-2)^3$ we are looking for a total of three vectors corresponding to $\lambda = 2$. First solve $(A-2)w = 0$. Specifically we solve the following matrix system.

$$\begin{bmatrix} 0 & 0 & 0 \\ 4 & 4 & 1 \\ -16 & -16 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$, which row-reduces to the system:

$$\begin{bmatrix} 4 & 4 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$, equivalent to $z = -4x-4y$, for any x, y. Thus

two independent solutions are $\alpha_1 = (1,0,-4)$, $\beta_1 = (0,1,-4)$.
Remember these are a basis of eigenvectors, i.e. a basis of $\ker(A-2)^1$.
Now consider the equation $(A-2)^2w = 0$. Since $(A-2)^2 = p(A) = 0$, any third vector which is independent of the first two will complete our basis of M_2. There is a method for finding a vector independent of two given ones, but $\alpha_2 = (0,0,1)$ clearly works. Do you see why?
Now let $a_2 = \alpha_2 = (0,0,1)$, and set $a_1 = (A-2)a_2 = (0,1,-4)$. Next we

have to eliminate either α_1 or β_1 replacing it by a_1. To do this, place these vectors $\{a_1,\alpha_1,\beta_1\}$ in the columns of a matrix, in that order, and row reduce. Then take the ones that correspond to the resulting "pivot" columns. I.e. if after reducing, the 2nd column is a pivot column, take the second vector above, α_1. If the third column is a pivot, take β_1 instead. Since the first two columns in the matrix:

$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ -4 & -4 & -4 \end{bmatrix}$, are obviously independent, there is no need to row reduce. Hence we let $b_1 = \alpha_1 = (1,0,-4)$. Thus an "upper" Jordan basis is $\{a_1,a_2,b_1\} = \{(0,1,-4), (0,0,1), (1,0,-4)\}$. Let's test it out:

$Aa_1 = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 6 & 1 \\ -16 & -16 & -2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ -8 \end{bmatrix} = 2\cdot a_1$, so indeed the first

column of the jordan matrix is (2,0,0). Then we have:

$Aa_2 = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 6 & 1 \\ -16 & -16 & -2 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} = a_1 + 2a_2$, so the second column of J is

(1,2,0). Continuing,

$Ab_1 = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 6 & 1 \\ -16 & -16 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -8 \end{bmatrix} = 2b_1$, and the 3rd column of J is

(0,0,2). Thus the (upper) Jordan matrix is indeed:

$J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. [Testing is wise; I worked this wrong, twice.]

Finally, if Q = the matrix with the Jordan basis as columns:

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ -4 & 1 & -4 \end{bmatrix}, \text{ then } Q^{-1}AQ = J.$$

Algorithm: It is somewhat tedious to do so, but we can now try to give a description of this procedure for finding Jordan bases:
(i) Find the characteristic polynomial $\chi = \Pi_\lambda (X-\lambda)^{m_\lambda}$. Then we know the space M_λ has dimension m_λ, and so we look for a total of m_λ independent vectors associated to λ.
(ii) Solve the system $(A-\lambda)w = 0$. There will be between one and m_λ independent solutions. If there are m_λ independent solutions, we have finished with the λ part of the basis. I.e. any choice of m_λ solutions $\{\alpha_1, \beta_1, \ldots, \gamma_1\} = \{a_1, b_1, \ldots, c_1\}$gives the λ part of the Jordan basis. The λ part of the matrix is diagonal with no "1's" above the diagonal.

$$\begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix}$$

(iii) If there is only one independent solution α_1 of $(A-\lambda)w = 0$, then there is only one λ block, of size $m_\lambda \times m_\lambda$. The equation $(A-\lambda)^2 w = 0$ will then have two independent solutions, of which α_1 is one, so we find another independent solution α_2. Continue in this way with α_3 another independent solution of $(A-\lambda)^3 w = 0$, ...etc. Eventually we find a basis $\{\alpha_1, \ldots, \alpha_{m_\lambda}\}$, for M_λ but not a cyclic one. We keep only the last vector found and change all the rest as follows. [We will write $m(\lambda)$ instead of m_λ.] First put $a_{m(\lambda)} = \alpha_{m(\lambda)}$. Then set $a_{m(\lambda)-1} = (A-\lambda)\cdot a_{m(\lambda)}$, and then $a_{m(\lambda)-2} = (A-\lambda)a_{m(\lambda)-1}, \ldots\ldots, a_2 = (A-\lambda)a_3$, $a_1 = (A-\lambda)a_2$. Then $\{a_1, \ldots, a_{m(\lambda)}\}$ is a cyclic upper Jordan basis for the λ part of the matrix, which is a single $m_\lambda \times m_\lambda$ block.

$$\begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}$$

(iv) If there are between 1 and m_λ independent solutions of $(A-\lambda)w = 0$, find a basis for them $\{\alpha_1,\beta_1,\ldots,\varepsilon_1\}$. the number of elements in this basis is the number of elementary Jordan blocks corresponding to λ. Then consider $(A-\lambda)^2 w = 0$, which is solved by all the vectors already found and also by at least one more. Enlarge the independent set $\{\alpha_1,\beta_1,\ldots,\varepsilon_1\}$ to a basis $\{\alpha_1,\beta_1,\ldots,\varepsilon_1, \alpha_2,\beta_2,\ldots,\delta_2\}$ of $\ker(A-\lambda)^2$ If there are now m_λ vectors altogether, stop, since we now have a basis for M_λ. If there are not m_λ vectors yet, then consider $(A-\lambda)^3 w = 0$. We will be able to find at least one more vector solving this equation, in addition to the ones already found. Enlarge the previous set to $\{\alpha_1,\beta_1,\ldots,\varepsilon_1, \alpha_2,\beta_2,\ldots,\delta_2, \alpha_3,\beta_3\ldots,\gamma_3\}$, a basis of $\ker(A-\lambda)^3$. Continue until there are m_λ vectors altogether, and we then have a basis of M_λ which we must change into a cyclic basis. For simplicity assume $\{\alpha_1,\beta_1,\ldots,\varepsilon_1, \alpha_2,\beta_2,\ldots,\delta_2, \alpha_3,\beta_3\ldots,\gamma_3\}$, is already a basis of M_λ, i.e. assume the minimal polynomial of A on M_λ is $(X-\lambda)^3$. (Notice that at each successive stage, the number of additional vectors found is less than or equal to the number found at the previous stage. I.e. the number of vectors with subscript "2" is no greater than the number with subscript "1", the number with subscript "3" is no greater than the number with subscript "2",.....etc.

(v) Take the last set of vectors found in step (iv), namely those with subscript 3, $\{\alpha_3,\beta_3\ldots,\gamma_3\}$, and rename them $\{a_3,b_3\ldots,c_3\}$. Now apply $(A-\lambda)$ to this set obtaining new vectors $\{a_2,b_2\ldots,c_2\}$. These must replace some of the vectors $\{\alpha_2,\beta_2,\ldots,\delta_2\}$ but we don't know yet which ones.

So consider the set $\{\alpha_1,\beta_1,\ldots,\varepsilon_1;a_2,b_2\ldots,c_2, \alpha_2,\beta_2,\ldots,\delta_2\}$ which generates $\ker(A-\lambda)^2$, and reduce it to an independent set, working from left to right. The set is already independent from the left end as far as c_2 [why?], hence you cannot lose any vectors until you get into the subset $\{\alpha_2,\beta_2,\ldots,\delta_2\}$. After this reduction, we have a basis $\{\alpha_1,\beta_1,\ldots,\varepsilon_1, a_2,b_2\ldots,c_2, \ldots,d_2; a_3,b_3,\ldots,c_3\}$, of M_λ, where the last part (the part named by Roman letters) is cyclic, but not the rest.

(vi) Now apply $(A-\lambda)$ to the set $\{a_2,b_2\ldots,c_2, \ldots,d_2\}$ obtaining the set $\{a_1,b_1\ldots,c_1, \ldots,d_1\}$ These will replace some of the vectors $\{\alpha_1,\beta_1,\ldots,\varepsilon_1\}$ To see which ones, form the set $\{a_1,b_1\ldots,c_1, \ldots,d_1,\alpha_1,\beta_1,\ldots,\varepsilon_1\}$, and again eliminate dependent vectors, working from left to right. This leaves a basis for $\ker(A-\lambda)$ of form

$\{a_1, b_1 \ldots, c_1, \ldots, e_1\}$.

(vi) Now the set $\{a_1, b_1, \ldots, e_1; a_2, b_2 \ldots, c_2, \ldots, d_2; a_3, b_3, \ldots, c_3\}$, is a cyclic basis for M_λ, but is not yet in the right order. Reorder it as follows: $\{a_1, a_2, a_3; b_1, b_2, b_3; \ldots; c_1, c_2, c_3; \ldots; d_1, d_2; \ldots, e_1\}$ i.e. first list all the blocks of maximal length (here they are of length three), then all the second longest blocks (here of length two), ...,finally all the blocks of length one. This puts the (λ-portion of the) basis in upper Jordan form. If you prefer lower Jordan form, reverse the order of this portion of the basis.

(vii) Repeat the whole procedure for the next characteristic root, i.e. the next value of λ. Finally place all the different λ - bases together, one after another. That gives the full Jordan basis. If Q is the matrix with this Jordan basis as its columns, then $Q^{-1}AQ = J$ is the Jordan form of A.

(viii) If the largest elementary blocks are larger than size three, just extend the procedure described in parts (iv)-(vi).

Let's work out another example from the previous section.

$$B = \begin{bmatrix} 0 & 1 & -2 & 1 \\ -2 & 1 & -6 & 3 \\ 2 & -3 & 0 & 1 \\ 2 & -3 & -2 & 3 \end{bmatrix}, \text{ where } \chi = (X-2)^3(X+2)$$

Consider $(B-2)w = 0$, i.e solve the system:

$$\begin{bmatrix} -2 & 1 & -2 & 1 \\ -2 & -1 & -6 & 3 \\ 2 & -3 & -2 & 1 \\ 2 & -3 & -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \text{ Row-reduction yields:}$$

$$\begin{bmatrix} 2 & -1 & 2 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \text{ which is equivalent to}$$

$y = -2z+u$, $2x = y-2z+u = 2y$, for any z,u. Thus a basis of $\ker(B-2)$ is given by $\{\alpha_1 = (-2, -2, 1, 0), \beta_1 = (1, 1, 0, 1)\}$. This tells

us there are two blocks corresponding to $\lambda = 2$. Hence one is 1×1 and the other is 2×2, and we need only one more basis vector for the space M_2. We solve next $(B-2)^2 w = 0$, i.e. since $(B-2)^2 =$

$$\begin{bmatrix} -2 & 1 & -2 & 1 \\ -2 & -1 & -6 & 3 \\ 2 & -3 & -2 & 1 \\ 2 & -3 & -2 & 1 \end{bmatrix}\begin{bmatrix} -2 & 1 & -2 & 1 \\ -2 & -1 & -6 & 3 \\ 2 & -3 & -2 & 1 \\ 2 & -3 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 8 & 16 & -8 \\ 0 & 8 & 16 & -8 \\ 0 & 8 & 16 & -8 \end{bmatrix}$$, reducing yields:

$$\begin{bmatrix} 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$, equivalent to $y = -2z + u$, for any z,u

We must enlarge our previous set $\{\alpha_1,\beta_1\}$ to a basis for this solution set. Looking at the old equations we need only choose y,z,u as before, and x differently. Eg. take $\alpha_2 = (0, 1, 0, 1)$.
Now we have a basis $\{\alpha_1,\beta_1,\alpha_2\}$ of M_2, but not a cyclic one. So put $a_2 = \alpha_2 = (0,1,0,1)$. Then set $a_1 = (B-2)a_2 = (2,2,-2,-2)$, and note this is another element of ker(B-2). Now recall we want to find out which of our previous elements in ker(B-2) to replace by this one. Consider the sequence $\{a_1,\alpha_1,\beta_1\} = \{(2,2,-2,-2), (-2,-2,1,0), (1,1,0,1)\}$, and reduce from left to right. Clearly the first two are independent, so take $b_1 = \alpha_1 = (-2,-2,1,0)$. Thus a cyclic basis of M_2 is: $\{a_1,a_2,b_1\}$ = $\{(2,2,-2,-2), (0,1,0,1), (-2,-2,1,0)\}$.

Now proceed to M-2, which we know is one dimensional, so just find one non zero vector in ker(B+2). This means solve the system:

$$\begin{bmatrix} 2 & 1 & -2 & 1 \\ -2 & 3 & -6 & 3 \\ 2 & -3 & 2 & 1 \\ 2 & -3 & -2 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$, which is equivalent to: $$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$,

or $x = 0$, $y = z = u$. Thus (0,1,1,1) works
If we put the characteristic root $\lambda = -2$ first, and $\lambda = 2$ next, we have the Jordan basis $\{(0,1,1,1), (2,2,-2,-2), (0,1,0,1), (-2,-2,1,0)\}$, and the upper Jordan matrix is the following:

$J = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$. Again if $Q = \begin{bmatrix} 0 & 2 & 0 & -2 \\ 1 & 2 & 1 & -2 \\ 1 & -2 & 0 & 1 \\ 1 & -2 & 1 & 0 \end{bmatrix}$, then $Q^{-1}BQ = J$.

Exercise #146) Find matrices Q which put each of the following matrices in upper Jordan form over $\mathbb{C}$:

(i) $A = \begin{bmatrix} 0 & \cdot 1 \\ 1 & 0 \end{bmatrix}$. (ii) $B = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$. (iii) $C = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}$.

(ii) $D = \begin{bmatrix} 1 & -2 & -1 & 0 \\ 1 & 0 & -3 & 0 \\ -1 & -2 & 1 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}$. (ii) $E = \begin{bmatrix} 5 & -1 & -3 & 2 & -5 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & -2 \\ 0 & -1 & 0 & 3 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{bmatrix}$.

The Jordan decomposition of an endomorphism.
An endomorphism whose characteristic polynomial splits over k has a Jordan matrix which is not necessarily diagonal, but the Jordan matrix does have a diagonal submatrix. I.e. a Jordan matrix is composed of two parts, the diagonal part and the off-diagonal part, the "λ's" and the "1's" This decomposition of the Jordan matrix is actually intrinsic to the endomorphism. I.e for any endomorphism T such that χ_T splits, there is a unique decomposition $T = S + N$, into a diagonalizable endomorphism S plus a nilpotent endomorphism N, such that $SN = NS$. This decomposition is the intrinsic version of the Jordan form. We prove this next.

Definition: An endomorphism $T: M \to M$ is called "nilpotent" iff for some integer $n > 0$, $T^n = 0$.

Remark: An endomorphism of an m-dimensional k-vector space M is nilpotent iff the minimal polynomial μ is power of X, i.e. iff $\mu = X^r$ where $0 < r \leq m$, iff the characteristic polynomial $\chi = X^m$. Since a diagonalizable endomorphism has square free minimal polynomial, a

nilpotent endomorphism T is diagonalizable iff $\mu = X$, iff $T = 0$.

Exercise #147): Prove every "strictly upper diagonal" square matrix is nilpotent. I.e if an $m \times m$ matrix A has zeroes on and below the diagonal, then $A^m = 0$. [Of course the same holds for strictly lower diagonal matrices.]

Now we are ready to prove our main result:
Theorem: If $T:M \to M$ is an endomorphism of a finite dimensional k-vector space, such that χ factors into linear factors over k, then T is uniquely expressible as a sum $T = S + N$, where S is diagonalizable, N is nilpotent, and where $NS = SN$. [For reasons which will appear in the next section, the diagonalizable part S of T is also called the "semi-simple" part.]
proof: [**Existence**] It suffices to express the (upper) Jordan matrix J_T as such a sum. We have only to split J_T into its diagonal entries and its off diagonal entries. I.e. let [S] be the diagonal matrix obtained from J_T by replacing all off-diagonal entries by zero, and let [N] be the strictly upper diagonal matrix obtained from J_T by replacing all diagonal entries by zero. Then $J_T = [S]+[N]$, and [S] is diagonal while [N] is nilpotent. Indeed, if r is the size of the largest elementary Jordan block of J_T, then $[N]^r = 0$. Since [S] is diagonal it represents a diagonalizable endomorphism $S:M \to M$, and since [N] is nilpotent, it represents a nilpotent endomorphism $N:M \to M$. Moreover since $J_T = [S]+[N]$, then $T = S+N$. To see that $SN = NS$, it suffices to show that $SNv = NSv$ for some basis v of M. We choose the Jordan basis $\{v_1, ..., v_m\}$ associated to J_T. On each v in this basis, S acts like multiplication by a scalar, and scalar multiplication commutes with every endomorphism
It is instructive to give an intrinsic description of this decomposition: If the minimal polynomial $\mu_T = \prod_\lambda (X-\lambda)^{r_\lambda}$, then the primary decomposition of M is $M \cong \prod_\lambda M_\lambda$ where $M_\lambda = \ker(T-\lambda)^{r_\lambda}$ is the "generalized λ-eigenspace" of T. [The usual λ-eigenspace is of course $\ker(T-\lambda)$.] The diagonalizable part S of T is the endomorphism which equals λI on M_λ. Then $N = T-S$; in particular $N = T-\lambda I$ on M_λ. Since each M_λ is a (usual) λ-eigenspace for S, S is diagonalizable. Since $N = T-\lambda I$ on M_λ, the minimal polynomial of T on M_λ is X^{r_λ}. Hence N is nilpotent on each M_λ, and thus N is nilpotent on M. The argument that $SN = NS$ is the same as above.

[Uniqueness] (This proof, by members of the class, originated in Patricio's idea to use the uniqueness of the Jordan form for T.)
Lemma: If $T = \tilde{S} + \tilde{N}$, where $\tilde{S}\tilde{N} = \tilde{N}\tilde{S}$, $\tilde{S}$ is diagonalizable, and $\tilde{N}$ is nilpotent, then there is a basis of M which is a Jordan basis simultaneously for $\tilde{N}$, for $\tilde{S}$, and for T.
proof: First note that if $\tilde{M}_\lambda$ is an eigenspace for $\tilde{S}$, then $\tilde{N}(\tilde{M}_\lambda) \subset \tilde{M}_\lambda$. I.e. if v is in $\tilde{M}_\lambda$ so that $\tilde{S}v = \lambda v$, then $\tilde{S}(\tilde{N}v) = \tilde{N}(\tilde{S}v) = \tilde{N}(\lambda v) = \lambda(\tilde{N}v)$, so $\tilde{N}v$ is again in $\tilde{M}_\lambda$. Since $\tilde{N}$ is nilpotent on all of M it is also nilpotent on $\tilde{M}_\lambda$, and thus its minimal polynomial on $\tilde{M}_\lambda$ is a power of X, hence splits in k. Thus there is a basis of $\tilde{M}_\lambda$ which is a Jordan basis for $\tilde{N}$. Since every vector in $\tilde{M}_\lambda$ is an eigenvector for $\tilde{S}$, this basis is also an eigenbasis, hence in particular a Jordan basis, for $\tilde{S}$ on M_λ. Since $\tilde{N}$ is nilpotent on $\tilde{M}_\lambda$, its Jordan matrix in this basis has zeroes on the diagonal, while the matrix of $\tilde{S}$ has λ along the diagonal and zeroes elsewhere. Hence adding the matrices for $\tilde{S}$ and for $\tilde{N}$ gives a matrix for T on $\tilde{M}_\lambda$ which is in Jordan form. Combining the bases obtained this way for each $\tilde{M}_\lambda$, gives a basis for M which is a Jordan basis for $\tilde{S}$, $\tilde{N}$ and T, as claimed. QED **Lemma.**

It follows from the lemma that for any decomposition of T into a sum $T = \tilde{S}+\tilde{N}$, with $\tilde{S}$ diagonalizable, $\tilde{N}$ nilpotent, and $\tilde{S}\tilde{N} = \tilde{N}\tilde{S}$, that $\tilde{S}$ and $\tilde{N}$ are respectively the diagonal and the off diagonal endomorphisms associated to some Jordan matrix of T. Then by the intrinsic description above of the Jordan decomposition, $\tilde{S} = \lambda I$ on each generalized eigenspace M_λ of T. Hence $\tilde{S} = S$, and $\tilde{N} = T-\tilde{S} = T-S = N$ also. In particular, T, S and $\tilde{S}$ all have the same characteristic polynomial, and the same characteristic roots λ, and for each λ we have $\tilde{M}_\lambda = M_\lambda$. QED **Theorem.**

Exercise #148) **a)** Show if $T = S+N$, where S,N are constructed from the Jordan matrix as above, then both S,N belong to the commutative ring k[T]. [Hint: It suffices to show that S belongs to the ring k[T]. If $\mu_T = \Pi_j(X-\lambda_j)^{r_i}$, and $k[X] \to \Pi_j k[X]/((X-\lambda_j)^{r_i})$ is the natural map, and f(X) maps to $([\lambda_1],\dots,[\lambda_n])$, then $f(T) = S$.]
b) If $T = \tilde{S}+\tilde{N}$ with $\tilde{S}$ diagonalizable, $\tilde{N}$ nilpotent, and $\tilde{S}\tilde{N} = \tilde{N}\tilde{S}$, show $\tilde{S},\tilde{N}$ commute with T and with the S,N in part a). Then show $S-\tilde{S}$ is diagonalizable, and $\tilde{N}-N$ is nilpotent. Deduce $\tilde{S} = S$, $\tilde{N} = N$, hence giving another proof of uniqueness of the jordan decomposition.

Remark: With hindsight we can see that we could have deduced the existence of the jordan form and the jordan decomposition at the same time. I.e. given the minimal polynomial of T, $\mu = \Pi_\lambda (X-\lambda)^{r_\lambda}$, we could decompose the space M by the primary decomposition into generalized eigenspaces $M \cong \Pi_\lambda M_\lambda$, where $M_\lambda = \ker(T-\lambda)^{r_\lambda}$. Then we can define the map S to be equal to λI on M_λ, and then define N = T-S. Since on M_λ, $N = (T-\lambda)$ has minimal polynomial X^r, it follows that $N:M_\lambda \to M_\lambda$ is nilpotent. Then following the steps in our construction of a jordan basis for N on M_λ, yields a basis whose associated matrix for N is in Jordan form (with zeroes on the diagonal) Since every basis for M_λ puts S in diagonal form, this same basis puts the matrix of T = S+N in Jordan form. Carrying this out for all M_λ, gives a jordan basis for T. With this approach, not using the cyclic decomposition theorem, we must prove uniqueness by some other method, such as the one outlined in exercise 137 above.

§11) Semi-simple endomorphisms, and "spectral theorems"
Since a diagonal matrix is the simplest, most useful type of matrix, it is helpful to have criteria guaranteeing that an endomorphism is diagonalizable, i.e. that the Jordan form is diagonal. We will prove several such results called "spectral theorems" in this section. Note that it follows immediately from the definition of the matrix associated to a basis, that $T:M \to M$ is diagonalizable iff there is a basis for M consisting of eigenvectors of T, i.e. iff there is a basis $\{v_1,\dots,v_m\}$ and scalars λ_j in k, such that for each j, $Tv_j = \lambda_j v_j$ The following criterion follows from the theory of Jordan forms, but it is instructive to give a direct proof.

Lemma: $T:M \to M$ is diagonalizable iff the minimal polynomial of T factors over k into distinct linear factors.
proof: Suppose the minimal polynomial μ of T factors as follows, $\mu = \Pi(X-\lambda_i)$, with all λ_i distinct. It follows from the primary decomposition lemma that if $M_i = \ker(T-\lambda_i)$, then $M \cong \Pi_i M_i$. Since by definition, each M_i has a basis of eigenvectors of T, the union of these bases is a basis for M consisting of eigenvectors for T. Hence T is diagonalizable
Conversely, if $\{v_1,\dots,v_m\}$ is an eigenbasis for M, $\lambda_1,\dots,\lambda_s$ are the distinct eigenvalues, and $M_i = \ker(T-\lambda_i)$ = {the subspace of M

spanned by the eigenvectors in the basis corresponding to λ_i), then $M \cong \Pi_i M_i$. [The "addition" map $\psi:\Pi_j M_i \to M$, $\psi(x_1,\ldots,x_s) = \Sigma x_i$, is injective since the set $\{v_1,\ldots,v_m\}$ is independent, and surjective since the set spans M.] Now let $\mu(X) = \Pi_{i=1,\ldots s}(X-\lambda_i)$. Since these factors commute, for each j we can write $\mu(T) = g_j(T)(T-\lambda_j)$, for some polynomial g_j. That is, we can apply $(T-\lambda_j)$ first if we choose. Since $(T-\lambda_j)$ annihilates M_j, then $g_j(T)(T-\lambda_j) = \mu(T)$ annihilates M_j for every j, hence $\mu(T)$ annihilates M. Thus the minimal polynomial of T divides μ, and since μ has distinct linear factors, so does the minimal polynomial of T. [Since M cannot be the product of a proper subset of the factors M_j, it follows from the first part of the proof that in fact $\mu = \Pi_{j=1,\ldots s}(X-\lambda_j)$.] QED.

The previous lemma is an important theoretical criterion, but in practice it is not always obvious what the minimal polynomial of a matrix is. Of course we know how to compute it, by diagonalizing the characteristic matrix (over k[X]), so the question of whether a given matrix is diagonalizable (over k) is in principle decidable, assuming we can factor the polynomials that occur in the standard matrix representation for the pair (M,T). Another approach is to compute the characteristic polynomial, factor it, and find a basis of $\ker(T-\lambda)$ for each characteristic root λ. Then T is diagonalizable iff together these eigenbases form a basis for M, i.e. iff there are altogether dim(M) independent eigenvectors. These criteria for diagonalizability amount essentially to saying: try to diagonalize T (by computing the Jordan form); if you succeed then T was diagonalizable. In some situations it makes sense to proceed that way. But if for example we had a matrix measuring 100×100, it could be challenging either to to find the minimal polynomial by diagonalizing the characteristic matrix, or to compute the determinant defining the characteristic polynomial. What we are looking for instead is a test that will guarantee in advance the matrix is diagonalizable. This is useful even if you later carry out the computations, since if they don't work out, you know you made a mistake. In cases where the computations are too large to actually carry out, these a priori criteria provide the only hope of knowing whether the matrix is diagonalizable.

Suppose we want to show a particular T is diagonalizable over k, by

proving an eigenbasis exists. Assume χ splits over k. This is a necessary condition by the previous lemma, and when k is algebraically closed it requires no calculation to verify. Then we get at least one eigenvector of T as follows from the next two lemmas:

Lemma: An endomorphism $T:M \to M$ of a finite dimensional vector space M is an isomorphism iff T is injective, iff T is surjective, iff $\ker(T) = \{0\}$, iff $\det(T) \neq 0$.
proof: If T is an isomorphism then it has a two sided inverse which in particular is an inverse as a set map. Hence T is both injective and surjective. If T is injective then certainly $\ker(T) = \{0\}$. Conversely if $\ker(T) = \{0\}$, then $Tv = Tw$ implies $T(v-w) = 0$, which implies $v-w$ is in $\ker(T) = \{0\}$. Thus $v=w$ and T is injective. If T is injective, and $\{v_1,\ldots,v_m\}$ is a basis of M, then we claim $\{Tv_1,\ldots,Tv_m\}$ must be independent. For if $\Sigma a_i(Tv_i) = 0$ is a dependency relation, then by linearity $T(\Sigma a_i v_i) = 0$, hence T injective implies $\Sigma a_i v_i = 0$, and since $\{v_i\}$ are independent all $a_i = 0$. Thus $\Sigma a_i(Tv_i) = 0$ was the trivial relation, and the set $\{Tv_i\}$ is independent. Now since more than m vectors in M must be dependent, every vector in M depends on the set $\{Tv_i\}$ which is thus a basis of M. It follows that T is surjective, since if w is in M, then we can write $w = \Sigma a_i Tv_i = T(\Sigma a_i v_i)$. Thus every w in M is in the image of T. If T is surjective then for any basis $\{v_i\}$ of M, the set $\{Tv_i\}$ generates M. Then this set could be reduced to a basis by removing vectors which depend on previous ones. Since on the other hand every basis of M has exactly m vectors in it, there are no dependent vectors in the set $\{Tv_i\}$, so by an argument reversing that above, T is also injective. The equivalence of these properties with $\det(T) \neq 0$, is assumed from the theory of determinants, (covered in the appendix). QED.

Lemma: An element λ of k is a root of the characteristic polynomial χ of T, iff T has an eigenvector v corresponding to λ.
proof: By definition, λ is a root of χ iff $\det(\lambda \cdot I - T) = 0$, iff $(\lambda \cdot I - T)$ is not invertible, iff $\ker(\lambda \cdot I - T) \neq \{0\}$, iff $\lambda v = Tv$ for some $v \neq 0$. QED.

So here is one approach to proving a map T is diagonalizable:
Step One: Prove χ factors into linear factors over k, true for example if $k = \bar{k}$. Then T has at least one eigenvector v_1 by the previous lemma.

Step Two. Try to use induction to complete $\{v_1\}$ to an eigenbasis for M, by finding an eigenbasis for the "rest of the space". In this step we try to restrict T to an endomorphism on a lower dimensional subspace $N \subset M$, such that $M \cong N \times \{k \cdot v\}$. Unfortunately there may not be such a subspace, and then we are in trouble. The spectral theorems consist of various conditions guaranteeing such subspaces exist, so that we can complete the proof. Let's be more precise:

Definition: **(i)**: If $T:M \to M$ is an endomorphism, a subspace $N \subset M$ is called "T-invariant" iff $T(N) \subset N$.
(ii): $T:M \to M$ is "semisimple" iff for every T-invariant subspace $N \subset M$, there is some other T-invariant subspace $L \subset M$ such that the natural addition $N \times L \to M$ [i.e. the one taking (x,y) to $x+y$], is an isomorphism.

Terminology: Two subspaces $L,N \subset M$ are said to be "complementary" iff the addition map $N \times L \to M$ is an isomorphism.

Exercise #149): Two subspaces $N, L \subset M$ are complementary iff the union of a basis for N with a basis for L gives a basis for M.

Exercise #150): Prove the map $A:k^2 \to k^2$ defined by the matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, is not semisimple, by proving the A-invariant subspace spanned by e_1 has no A-invariant complement.

[**Definition**: In general, an R module Z is called semisimple if every submodule $X \subset Z$ is a direct factor of Z. I.e. iff for every R-submodule $X \subset Z$, there is some other R-submodule $Y \subset Z$ such that the natural map $X \times Y \to Z$ is an isomorphism. [Equivalently, an exact sequence $0 \to X \to Z \to W \to 0$, with Z in the middle, always splits.]

Examples: If k is any field, every finite dimensional k-vector space V is k-semisimple, since a basis for a subspace extends to a basis for V. The integers $\mathbf{Z}$ do not form a semisimple $\mathbf{Z}$-module, since the submodule $2\mathbf{Z} \subset \mathbf{Z}$ does not split off as a direct factor.

Remark: A k- endomorphism $T:M \to M$ is semi simple iff the pair (M,T) is a semisimple $k[X]$ module, iff M is a semisimple $k[T]$ module.] Semisimplicity is exactly the property we need to characterize

diagonalizable endomorphisms:

Lemma: An endomorphism $T:M \to M$ of a finite dimensional vector space over k, is diagonalizable iff the characteristic polynomial χ splits in k, and T is semisimple.

proof: [if]: Assume T is semisimple and χ factors over k into linear factors. If χ has a root λ_1 in k, then T has an eigenvector v_1, by the lemma above. The subspace N_1 spanned by v_1 is T-invariant, hence by hypothesis there is a complementary T-invariant subspace $P_1 \subset M$ such that $M \cong N_1 \times P_1$. If we consider the restriction $T_1: P_1 \to P_1$, of T to P_1, and choose a basis $\{w_2,\ldots,w_m\}$ of P_1, then with respect to the basis $\{v_1,w_2,\ldots,w_m\}$ of M, the matrix of T is a block matrix, with λ_1 in the (1,1) position as a 1×1 block, and the matrix of the restriction T_1 in the lower right hand corner as an (m-1)×(m-1) block:

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & & & & \\ 0 & & T_1 & & \\ \vdots & & & & \\ 0 & & & & \end{bmatrix}$$

Hence $\det(X \cdot I - T) = (X - \lambda_1) \cdot \det(X \cdot I - T_1)$.

Thus the characteristic polynomial χ_1 of T_1 divides the characteristic polynomial χ of T, hence χ_1 also factors over k into linear factors. Thus there exists an eigenvector v_2 for T_1, which is also an eigenvector for T. Then v_1 and v_2 are independent and the subspace N_2 spanned by v_1, v_2 is T- invariant, and thus has a T-invariant complement P_2. The same argument with block matrices shows that the restriction T_2, of T to P_2, also has an eigenvector v_3. Continuing in this way, we get an eigenbasis of M. Hence T is diagonalizable

[only if]: Assume T is diagonalizable, with characteristic polynomial χ. If we calculate χ from a diagonal matrix for T, it is obvious that $\det(X-T)$ is a product of linear factors. Now let $N \subset M$ be a T-invariant subspace, and let $\{w_1, \ldots, w_r\}$ be a basis for N. We want to find a T-invariant complement for N. Note that any subspace spanned by eigenvectors is T-invariant. Since T is diagonalizable, there is some basis for M consisting of eigenvectors $\{v_1,\ldots,v_m\}$.

Arrange these two sets of vectors as follows:
$\{w_1,...,w_r; v_1,...,v_m\}$, and starting at the left end, eliminate any vector which depends on the vectors to its left, as we have done before. After this is over we are left with a basis for M, of form $\{w_1,...,w_r; v_\alpha,...,v_\gamma\}$, which starts out with the basis $\{w_1,...,w_r\}$ of N, and continues with some eigenvectors $\{v_\alpha,...,v_\gamma\}$ of T. Then the subspace P spanned by the eigenvectors $\{v_\alpha,...,v_\gamma\}$, is a T-invariant complement of N, so T is semisimple.
QED Prop.

Corollary: If M is a finite dimensional vector space over k, and $T:M \to M$ an endomorphism whose characteristic polynomial splits in k, for example if k is an algebraically closed field such as $\mathbb{C}$, then T is diagonalizable iff T is semisimple.

So how can we recognize that an endomorphism is semisimple? Think of an endomorphism T of $\mathbb{R}^3$ that maps the z-axis, for example, into itself. Is there a property that would insure that T also maps the (x,y) plane into itself? If so, then the (x,y) plane would be a complementary subspace to the z-axis. We might think of a rotation about the z-axis. This is a good example of a semisimple endomorphism. It always preserves angles, hence if v,w, are perpendicular, then so are Tv,Tw. This implies that if N is an invariant subspace, then so is the subspace of vectors perpendicular to N, and this provides an invariant complement. Unfortunately, the characteristic polynomial of a rotation does not factor into linear factors over $\mathbb{R}$. [A 90^0 rotation of $\mathbb{R}^3$ has characteristic polynomial $\chi = (X-1)(X^2+1)$.] So a rotation is semisimple but (usually) not diagonalizable over $\mathbb{R}$.

This idea of using "orthogonal complements" for showing an endomorphism is semisimple *is* nonetheless the central method used in all finite dimensional spectral theorems. We just translate the method into the setting of complex vector spaces, where we can also assume the characteristic polynomial splits into linear factors. Then as a bonus we also get one theorem that holds over $\mathbb{R}$.
Recall that "metric" concepts from geometry such as length, angles, and perpendicularity, are provided in vector space theory by means of "dot products", or "inner products". Over $\mathbb{C}$, we make a variation

in the definition of the inner product, in order to have lengths come out to be real numbers, as follows:

Definition: The "**standard hermitian product**" on $\mathbb{C}^m$ is defined as follows: If $z = (z_1,\ldots,z_m)$, $w = (w_1,\ldots,w_m)$ are vectors in $\mathbb{C}^m$, then we define $\langle z,w\rangle = \Sigma z_j\overline{w}_j$. We may also denote $\langle z,w\rangle$ by $z\cdot w$.

Remark: Recall that the "transpose" of an $m\times n$ matrix A is the $n\times m$ matrix A^t whose columns are the rows of A. Then $(AB)^t = B^tA^t$. If z,w are column vectors, then $\langle z,w\rangle = z^t\overline{w}$ where the multiplication is matrix multiplication, and z^t is the transpose of z.

Properties of $\langle z,w\rangle$ on $\mathbb{C}^m$:

(i) bi-additivity: $\langle z+u, w\rangle = \langle z,w\rangle + \langle u,w\rangle$,
and $\langle z,w+u\rangle = \langle z,w\rangle + \langle z,u\rangle$.

(ii) (sesqui)- linearity: for λ in $\mathbb{C}$, $\langle \lambda z,w\rangle = \lambda\langle z,w\rangle$, $\langle z,\lambda w\rangle = \overline{\lambda}\langle z,w\rangle$.

(iii) hermitian symmetry: $\langle z,w\rangle = \overline{\langle w,z\rangle}$.

(iv) positivity: $\langle z,z\rangle$ is real, and if $z\neq 0$, then $\langle z,z\rangle > 0$.

(v) compatibility: Under the real isomorphism $\mathbb{C}^m \cong \mathbb{R}^{2m}$, [where $z = (\ldots,z_j,\ldots)$ corresponds to $(\ldots,x_j,y_j,\ldots)$, with $z_j = x_j+y_j$], $\langle z,z\rangle = \Sigma_j(x_j^2+y_j^2)$ = the squared Euclidean length of the vector z.

Definition: We say z and w are **orthogonal** iff $z\cdot w = 0$, and we define the **length** of z to be $|z| = (z\cdot z)^{1/2} \geq 0$. Note that a vector is zero iff it has length zero.

We can define a "**hermitian product**" on any complex vector space M, to be a pairing $M\times M\to\mathbb{C}$, satisfying properties (i)-(iv). We may call such a space a "**hermitian space**". Every finite dimensional complex vector space has such products, since we can choose a basis $\{v_1,\ldots,v_m\}$, hence an isomorphism of M with $\mathbb{C}^m$, and define a product on M by composing the isomorphism with the standard product, i.e. by $M\times M\to\mathbb{C}^m\times\mathbb{C}^m\to\mathbb{C}$. If we define a hermitian product on M this way, then the basis vectors $\{v_j\}$ used in the isomorphism all have length one, and are all mutually orthogonal Such a basis is called "**orthonormal**". Conversely, if M has a hermitian product, and $\{v_j\}$ is an orthonormal basis of M, then the associated isomorphism $M\to\mathbb{C}^m$ carries the product on M into the

standard product on $\mathbb{C}^m$. But do orthonormal bases always exist?

Orthonormal bases are easy to find if $\dim_{\mathbb{C}}(M) = 2$, as follows: Suppose M has basis $\{v,w\}$. Then for every λ in $\mathbb{C}$, $\{v, w-\lambda v\}$ is also a basis. We claim that λ can be chosen so that the new basis vectors $\{v, w-\lambda v\}$ are at least mutually orthogonal. This is easy since we just need $(w-\lambda v)\cdot v = 0$, i.e. we need $(w\cdot v) - \lambda(v\cdot v) = 0$, which holds iff $\lambda = (w\cdot v)/(v\cdot v)$. Now that we have an orthogonal basis, dividing every basis vector by its length gives an orthonormal basis. In general, use the next exercise.

Exercise #151): Assume M has a hermitian product, and that $\{v_1,\dots,v_s, w_{s+1},\dots,w_m\}$ is a basis for M over $\mathbb{C}$, such that the first s basis vectors $\{v_1,\dots,v_s\}$ are mutually orthogonal. Prove one can choose $\lambda_1,\dots,\lambda_s$ so that if $v_{s+1} = w_{s+1}-\lambda_1 v_1-\dots-\lambda_s v_s$, then $\{v_1,\dots,v_s, v_{s+1}, w_{s+2},\dots,w_m\}$ is still a basis for M, and the first $s+1$ vectors $\{v_1,\dots,v_s, v_{s+1}\}$ are now mutually orthogonal. Deduce that given any basis $\{w_1,\dots,w_m\}$ of M, there is an orthonormal basis $\{v_1,\dots,v_m\}$ such that for every s, the sets $\{w_1,\dots,w_s\}$ and $\{v_1,\dots,v_s\}$ span the same subspace. In particular a finite dimensional hermitian space always has an orthonormal basis

Terminology: This procedure for changing any basis into an orthonormal one is often called the "Gram-Schmidt" process.

Remarks: (i) Since the Gram Schmidt process provides a rational formula, with non zero denominator, for each of the new basis vectors v_s in terms of the old ones $\{w_1,\dots,w_s\}$, it follows that if the old vectors depend continuously on some variable then the new ones depend continuously on that variable also. This gives one solution to the problem [cf Ex. #121] of showing that a pair of continuous vector fields σ,τ on the sphere S, which are independent at each point p of S, would yield an isomorphism $\mathcal{C}\times\mathcal{C}\cong\mathcal{V}$, between the free rank two module on the ring $\mathcal{C}$ of continuous functions on the sphere, and the module $\mathcal{V}$ of all continuous vector fields on S. I.e. the map $\Theta:\mathcal{C}\times\mathcal{C}\to\mathcal{V}$ taking $(f,g)\mapsto f\sigma+g\tau$ is well defined with image in $\mathcal{V}$, since when f,σ,g,τ are all continuous, the linear combination $f\sigma+g\tau$ is continuous as well. The difficulty is to show Θ is invertible; i.e. that for each v in $\mathcal{V}$, the unique functions f,g such that $v(p) =$

$f(p)\sigma(p)+g(p)\tau(p)$ for all p in S, are continuous. If however we define the vector field γ by $\gamma = \tau - (\langle\sigma,\tau\rangle/\langle\sigma,\sigma\rangle)\,\sigma$, then γ is continuous and (σ,γ) are orthogonal and non zero for all p. Consequently, $\tilde{\sigma} = \sigma/|\sigma|$, and $\tilde{\gamma} = \gamma/|\gamma|$, are orthonormal vector fields. Thus for any v in $\mathcal{V}$, we can solve the equation $v = f\tilde{\sigma}+g\tilde{\gamma}$ for f,g by setting $f = \langle v,\tilde{\sigma}\rangle$, and $g = \langle g,\tilde{\gamma}\rangle$, which shows that f,g, are continuous. Thus the map $\mathcal{C}\times\mathcal{C}\to\mathcal{V}$ defined by $(f,g)\mapsto f\tilde{\sigma}+g\tilde{\gamma}$ would be invertible, implying $\mathcal{C}\times\mathcal{C}\cong\mathcal{V}$. [Of course, remember such vector fields σ,τ do not exist, and in fact $\mathcal{C}\times\mathcal{C}$ and $\mathcal{V}$ are *not* isomorphic.]

(ii) Gang Yu gave another solution of Ex. #121 by showing how to solve $v = f\sigma+g\tau$ directly for the functions f,g as follows: instead of needing σ,τ to be perpendicular in order to eliminate one of them by using the dot product, he observed that the cross product will eliminate one of them even without perpendicularity. I.e. since $\tau\times\tau = 0 = \sigma\times\sigma$, and $\sigma\times\tau\neq 0$, if $v = f\sigma+g\tau$, then $v\times\tau = f(\sigma\times\tau)$, and $\sigma\times v = g(\sigma\times\tau)$, hence $f = (v\times\tau)\cdot(\sigma\times\tau)/(\sigma\times\tau)\cdot(\sigma\times\tau)$, and $g = (\sigma\times v)\cdot(\sigma\times\tau)/(\sigma\times\tau)\cdot(\sigma\times\tau)$.

Exercise #152): Assume M has a hermitian product. Prove that if $(v_j)_{j=1....m}$ is an orthonormal basis for M over $\mathbb{C}$, and we define $\varphi: M\to\mathbb{C}^m$ by $\varphi(\Sigma a_j v_j) = (a_1,....,a_m)$, then φ carries the dot product on M into that on $\mathbb{C}^m$; i.e. $(\Sigma a_j v_j)\cdot(\Sigma b_j v_j) = \Sigma a_j\bar{b}_j$.

The value of these hermitian products in producing complementary subspaces is based on the concept of "orthogonal complements".

Definition: If M is a complex vector space with a hermitian product, and if $N\subset M$ is any subset, then define $N^{\perp} = \{w \text{ in } M : \langle v,w\rangle = 0 \text{ for all } v \text{ in } N\}$.

Terminology: For any subset $N\subset M$, the subset $N^{\perp}\subset M$ is called the **orthogonal complement** of N, (in M).

Exercise #153) If M is a finite dimensional complex hermitian space, and if $S\subset M$ is any subset, then $S^{\perp}$ is a subspace of M. If $N\subset M$ is a subspace, then the addition map $N\times N^{\perp}\to M$ is an isomorphism, i.e. $N^{\perp}$ is a complement to N.

To understand the relation between endomorphisms and inner

products, we need the concept of the "adjoint" of an endomorphism. The adjoint of T will be an endomorphism T^* such that $\langle Tz,w\rangle = \langle z,T^*w\rangle$ for all z,w in M. We need to show that such an endomorphism exists and is unique, at least in finite dimensions.

The easiest way to do this is to use an orthonormal basis to give an isomorphism $(M,\langle,\rangle) \cong (\mathbb{C}^m, \langle,\rangle)$, i.e. an isomorphism $M\cong\mathbb{C}^m$ that carries our hermitian product over into the usual product on $\mathbb{C}^m$. Then we can compute with matrices. I e. if v, w are column matrices in $\mathbb{C}^m$, then $\langle v,w\rangle = v^t\overline{w}$. Thus for any matrix A, we have $\langle Av,w\rangle = (Av)^t\,\overline{w}$. We want to find a matrix B such that $\langle Av,w\rangle = \langle v,Bw\rangle = v^t\,\overline{B}\overline{w}$. Since $(Av)^t\,\overline{w} = v^tA^t\overline{w}$, this means we want $v^tA^t\overline{w} = v^t\,\overline{B}\overline{w}$. Obviously this would be true if $\overline{B} = A^t$. I.e. just set $B = \overline{A}^t$. Thus if A is the matrix of an endomorphism T, with respect to an *orthonormal* basis, then the matrix of the adjoint T^*, is the "transpose conjugate" of the matrix for T Thus for a complex matrix A, "A^*" denotes $\overline{A}^t$, the adjoint of A with respect to the usual hermitian product on $\mathbb{C}^m$. Uniqueness follows from the general fact that if a vector is orthogonal to every vector, then it is orthogonal to itself, hence equals zero. I.e. suppose for every v, we had $\langle Av,w\rangle = \langle v,Bw\rangle = \langle v,Cw\rangle$. Then for every v, $\langle v,(B-C)w\rangle = 0$. Thus for $v = (B-C)w$, we have $\langle (B-C)w,(B-C)w\rangle = |(B-C)w| = 0$. Thus $(B-C)w = 0$. Since this holds for all w, B=C.

There is also an abstract way to produce the adjoint as an endomorphism, without using matrices, (but then you still need to know what the matrix of the adjoint is). Let's see that way, too.
Definition: For any complex vector space M, define the "dual space" $M^* = \mathrm{Hom}_{\mathbb{C}}(M,\mathbb{C})$.

Lemma: If M is an m-dimensional vector space over $\mathbb{C}$ with basis $\{v_j\}$, then the homomorphisms $\{\lambda_j\}$ defined by $\lambda_j(v_i) = 0$ if $i\neq j$, and $\lambda_j(v_j) = 1$, give a basis for M^*, called the dual basis to $\{v_j\}$. In particular, M^* is also m -dimensional over $\mathbb{C}$, hence $M\cong M^*$.
proof: The $\{\lambda_j\}$ are independent, since if $\Sigma a_j\lambda_j = 0$, then $0 = (\Sigma a_j\lambda_j)(v_i) = a_i\lambda_i(v_i) = a_i$, for all i. On the other hand, the $\{\lambda_j\}$ generate M^*, since if $\varphi: M\to\mathbb{C}$ is any homomorphism, and if $a_j = \varphi(v_j)$, then we claim $\varphi = \Sigma a_j\lambda_j$. To see this, evaluate both sides at

v_j. We get $\varphi(v_j) = a_j = (\Sigma a_j\lambda_j)(v_j)$. Since φ and $\Sigma a_j\lambda_j$ agree on a basis, they agree everywhere. QED.

Exercise #154): If M is a finite dimensional complex vector space with a hermitian product, prove that the map $\psi: M \to M^*$ such that $\psi(w) = \langle \cdot, w \rangle$ is a "conjugate-linear" isomorphism; [$\psi(\lambda w) = \bar{\lambda}\psi(w)$]. (Here the dot is just a place holder for the argument of the function, i.e. $\psi(w)(v) = \langle v,w \rangle$, for every v in M.)

Lemma: Given any endomorphism $T: M \to M$, there is a unique endomorphism $T^*: M \to M$ s.t. $\langle Tz,w \rangle = \langle z,T^*w \rangle$ for z,w in M. If $\{v_j\}$ is an orthonormal basis of M, then with respect to this basis, the matrix of T^* is the "conjugate transpose" of the matrix of T: i.e. the (i,j) entry of $[T^*]$ is the complex conjugate of the (j,i) entry of [T]
Proof: Given w in M, the map $\varphi: M \to \mathbb{C}$ defined by $\varphi(z) = \langle z,w \rangle$ belongs to M^* by the previous exercise. Composing with $T: M \to M$ gives $(\varphi \circ T): M \to \mathbb{C}$, also in M^*. Thus for z in M, $(\varphi \circ T)(z) = \langle Tz,w \rangle$. Again by the previous exercise, there is a unique element of M, call it T^*w, such that $(\varphi \circ T) = \langle \cdot, T^*w \rangle$. Thus for every v in M, $\langle Tv,w \rangle = \langle v,T^*w \rangle$. Now we must show that the map taking w to T^*w defines a homomorphism $T^*: M \to M$. To show $T^*(u+w) = T^*u + T^*w$, it suffices to show that both represent the same function $M \to \mathbb{C}$. I.e. in general if $\langle v,x \rangle = \langle v,y \rangle$ for all v, then $x = y$. [This is because then $\langle v,x-y \rangle = 0$ for all v including for $v = x-y$, whence $x-y = 0$.] Thus it suffices to show for every v in M, that $\langle v,T^*(u+w) \rangle = \langle v,T^*u \rangle + \langle v,T^*w \rangle$. But $\langle v,T^*u \rangle + \langle v,T^*w \rangle = \langle Tv,u \rangle + \langle Tv,w \rangle = \langle Tv,u+w \rangle = \langle v,T^*(u+w) \rangle$. Similarly, to show $T^*(\lambda w) = \lambda T^*(w)$, it suffices to show for all z that $\langle z,T^*(\lambda w) \rangle = \langle z,\lambda T^*(w) \rangle$. But $\langle z,T^*(\lambda w) \rangle = \langle Tz,\lambda w \rangle = \bar{\lambda}\langle Tz,w \rangle = \bar{\lambda}\langle z,T^*(w) \rangle = \langle z,\lambda T^*w \rangle$. Thus T^* is $\mathbb{C}$-linear.
After choosing an orthonormal basis of M, T and T^* are represented by matrices acting on $\mathbb{C}^m$ and the hermitian product on M is identified with the standard one on $\mathbb{C}^m$. If T is a matrix acting on $\mathbb{C}^m$, and $\bar{T}^t$ is its transpose conjugate, we want to show that $\langle z, \bar{T}^t w \rangle = \langle Tz,w \rangle$ for all z,w in $\mathbb{C}^m$. If $T = [a_{ij}]$, $z = (z_j)$, and $w = (w_i)$, then $T(z) = (\Sigma_j a_{1j}z_j, \ldots, \Sigma_j a_{mj}z_j)$, and $\bar{T}^t w = \Sigma_i \bar{a}_{ij}w_i$. Thus $\langle Tz,w \rangle = \Sigma_i(\Sigma_j a_{ij}z_j)\bar{w}_i) = \Sigma_j z_j(\Sigma_i a_{ij}\bar{w}_i) = \langle z,\bar{T}^t w \rangle$. QED.

Definition: The map T^* is called the (hermitian) "adjoint" of T.

Remark: If M is a complex hermitian space, and $T:M \to M$ an endomorphism, then $(T^*)^* = T$.

If an endomorphism is well behaved with respect to the hermitian product, then we can use an inductive argument, as outlined above, to produce not just an eigenbasis, but an orthonormal eigenbasis.

Definition: An endomorphism $T:M \to M$ of a (finite dimensional) hermitian space is called "**unitarily diagonalizable**" iff M has an *orthonormal* basis of eigenvectors for T

Definition: An endomorphism of a hermitian space such that $T = T^*$ is called **hermitian**, or **self adjoint**.

Exercise #155) (spectral theorem for hermitian operators): Prove if $T=T^*$ then T is unitarily diagonalizable. [Hint: The orthogonal complement of a T-invariant subspace is T-invariant.]

Remark: There are a lot of hermitian operators, since all we have to do to produce them on $\mathbb{C}^m$ is construct a matrix whose transpose equals its complex conjugate. In particular the entries above the diagonal can be any complex numbers and the diagonal entries can be any real numbers.

There is another similarly easy and standard diagonalization result:
Definition: An endomorphism $T:M \to M$ of a hermitian space is called **unitary** iff $T^* = T^{-1}$.

Exercise #156)(spectral theorem for unitary operators): **(i)** Show T is unitary *iff* $\langle v,w \rangle = \langle Tv,Tw \rangle$ for all v,w. **(ii)** Prove that a unitary operator is unitarily diagonalizable.

Definition: A complex *matrix* P is called **unitary** iff $P^{-1} = \bar{P}t$.

Exercise #157): Show a complex matrix P is unitary iff its columns form an orthonormal basis of $\mathbb{C}^m$, iff its rows also form an orthonormal basis.

Remark: An endomorphism $T:\mathbb{C}^m \to \mathbb{C}^m$ with matrix A, is unitarily diagonalizable (with respect to the usual hermitian product on $\mathbb{C}^m$) iff there is a unitary matrix U such that $U^{-1}AU$ is diagonal.

Definition: The unitary $m \times m$ matrices form a group, called the **unitary group**, denoted U(m). The subgroup of unitary matrices of determinant one, is called the **special unitary group**, SU(m).

Exercise #158): Tell how to determine whether two elements of the group U(m) are conjugate or not. What about in SU(m)?

There is a spectral theorem which generalizes the ones you have proved in the previous exercises, as follows:

Definition: A "normal" operator on a complex hermitian space M is an endomorphism $T:M \to M$ such that $TT^* = T^*T$, i.e. one that commutes with its adjoint.

Remark: Hermitian and unitary operators are special cases of normal operators

Normal operators are the largest class admitting orthonormal bases of eigenvectors, as we now show, (thus proving again Exs. 144, 145).

Spectral Theorem for Normal Operators:
Theorem: If $T:M \to M$ is an endomorphism of a finite dimensional hermitian space, then T is unitarily diagonalizable iff T is normal.
proof: Assuming T is unitarily diagonalizable, the orthonormal eigenbasis gives a product preserving isomorphism with $\mathbb{C}^m$ such that the matrix of T is diagonal. Since the matrix of the adjoint is the transpose conjugate of that of T, it is also diagonal. Hence $[T]$, and $[T^*]$ commute, and thus so do the operators T, T^*.
Conversely, If T is normal, let λ be any characteristic root and M_λ the corresponding eigenspace. Then we claim $T^*(M_\lambda) \subset M_\lambda$. This is the same as the proof in the theorem in the previous section on Jordan decompositions. I.e. if v is in M_λ, then $T(T^*v) = T^*(Tv) = T^*(\lambda v) = \lambda(T^*v)$. Thus T^*v is in M_λ. Next we claim that

$T(M_\lambda^\perp) \subset (M_\lambda^\perp)$. I.e. assume v is in M_λ and w is in $M_\lambda^\perp$, then T^*v is in M_λ, so $0 = \langle T^*v, w\rangle = \langle v, Tw\rangle$, hence Tw is in $(M_\lambda^\perp)$. Similarly, $T^*(M_\lambda^\perp) \subset (M_\lambda^\perp)$. In particular the adjoint of the restriction of T is the restriction of the adjoint. So consider the map $T:(M_\lambda^\perp) \to (M_\lambda^\perp)$. Since $TT^* = T^*T$ on all of M, the same is true on $(M_\lambda^\perp)$. Thus the restriction of T to $(M_\lambda^\perp)$ is normal, and by induction on the dimension of the hermitian space, $(M_\lambda^\perp)$ has an orthonormal basis of eigenvectors of T. Since we can find a basis of M_λ and then make it orthonormal, by joining these two bases, we get an orthonormal basis of M consisting of eigenvectors of T. QED.

Exercise #159) (i) If T is hermitian, $Tv = \lambda v$, $Tw = \mu w$, and $\lambda \neq \mu$, prove $\langle v, w\rangle = 0$. (Remember λ, μ are real.)
(ii) If T is unitary, $Tv = \lambda v$, $Tw = \mu w$, and $\lambda \neq \mu$, prove $\langle v, w\rangle = 0$. (Hint: first prove $|\lambda| = |\mu| = 1$.)

Exercise #160) Prove: (i) If S is normal, then for every v, $|Sv| = |S^*v|$. (ii) If T is normal, then $Tv = \lambda v$ iff $T^*v = \overline{\lambda} v$. [Hint: Show $|(\lambda - T)v| = 0$ iff $|(\overline{\lambda} - T^*)v| = 0$.]

Spectral theorem for real symmetric operators
A nice surprise is that we get a spectral theorem for real matrices using the result for hermitian ones. Note first the following:

Lemma: If $T:M \to M$ is a hermitian operator, and λ is a characteristic root of T, then λ is real.
proof: We know λ is also an eigenvalue of T so there exists $v \neq 0$ such that $Tv = \lambda v$. Then $\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle$. Since $v \neq 0$, then $\langle v, v\rangle \neq 0$, so $\lambda = \overline{\lambda}$, and λ is real. [We could also unitarily diagonalize T. Since the matrix of T is then self adjoint, the diagonal entries are self conjugate, hence real.] QED.

Corollary: The characteristic polynomial of a hermitian operator T has real coefficients and real roots.
proof: Compute it from a diagonal matrix for T. QED.

Definition: The standard dot product on $\mathbb{R}^m$ is given by $\langle x, y\rangle = x \cdot y = x^t y = \Sigma_i x_i y_i$, where $x = (x_i)$, and $y = (y_i)$.

Properties of $\langle x,y\rangle$ on $\mathbb{R}^m$:
(i) bi-additivity: $\langle x + t, y\rangle = \langle x,y\rangle + \langle t,y\rangle$, and $\langle x,t+y\rangle = \langle x,t\rangle + \langle x,y\rangle$.
(ii) bi-linearity: for λ in $\mathbb{R}$, $\langle \lambda x,y\rangle = \lambda\langle x,y\rangle = \langle x,\lambda y\rangle$.
(iii) symmetry: $\langle x,y\rangle = \langle y,x\rangle$.
(iv) positivity: $\langle x,x\rangle$ is real, and if $x \neq 0$, then $\langle x,x\rangle > 0$.

Definition: If x is in $\mathbb{R}^m$, then $|x| = \text{length}(x) = (x\cdot x)^{1/2}$.
We define an "inner product" on any real vector space M to be a pairing $M\times M\to\mathbb{R}$ satisfying the properties above. Such an M is called, imaginatively, an inner product space. Again any basis for a finite dimensional real vector space M defines an isomorphism $M\to\mathbb{R}^m$ which allows us to transfer the usual dot product from $\mathbb{R}^m$ to M. If M has a dot product and is finite dimensional, the Gram-Schmidt process gives an orthonormal basis for M which simultaneously identifies M with $\mathbb{R}^m$ and identifies the dot product on M with the usual one on $\mathbb{R}^m$. Given any endomorphism $T:M\to M$, there is a unique endomorphism $T^*:M\to M$ such that $\langle Tx,y\rangle = \langle x,T^*y\rangle$ for all x,y in M. If we identify M with $\mathbb{R}^m$ by choosing an orthonormal basis, then the associated matrix for T^* is the transpose of the metrix for T.
We could prove these facts exactly as we did for hermitian spaces, but you can easily do that, so for fun let's prove something different.

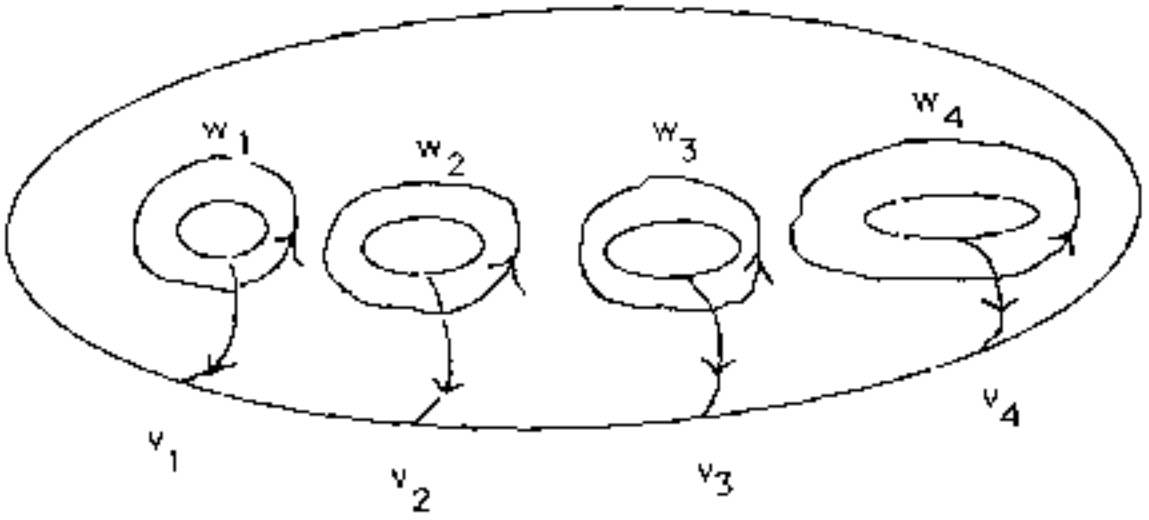
Definition: If M is a real vector space, the space $\text{Hom}_{\mathbb{R}}(M,\mathbb{R}) = M^*$ is the **dual space** of M.

Note there is always a natural pairing $M\times M^*\to\mathbb{R}$, the "evaluation" pairing, $\langle v,\varphi\rangle = \varphi(v)$. This is bi-additive and bilinear, but symmetry and positivity do not make sense. By analogy with the hermitian case, an inner product on M yields an isomorphism of a space with its dual, i.e. the map $M\to M^*$ taking w to $\langle\cdot,w\rangle$ is an isomorphism. But to what extent is the converse true?

Lemma: If M is a finite dimensional, real vector space and $\psi:M\to M^*$ is an isomorphism, we get a bi-additive, bilinear product on M by setting $\langle v,w\rangle = \psi(v)(w)$ This product is "non singular" in the sense that $\langle v,w\rangle = 0$ for all w iff $v = 0$, and $\langle v,w\rangle = 0$ for all v iff $w = 0$. Symmetry and positivity do not necessarily hold.

proof: The only slightly non trivial part is to show if $w \neq 0$ in M, then for some v, $\langle v,w\rangle \neq 0$. Since ψ is an isomorphism, it suffices to show there is homomorphism φ in M^* such that $\varphi(w) \neq 0$. Extend the set $\{w\}$ to a basis of M, and then we can define a map $\varphi: M \to \mathbb{R}$ by defining it any way we like on the basis, so let $\varphi(w) = 1$. QED.

Example: Let M have even real dimension $m = 2n$, with basis $\{v_1,\ldots,v_n,w_1,\ldots,w_n\}$. Let $\{v_1^*,\ldots,v_n^*,w_1^*,\ldots,w_n^*\}$ be the dual basis of M^*, i.e. $v_j^*(w_i) = 0$ for all i,j, $v_j^*(v_i) = 0$ if $i \neq j$, and $v_i^*(v_i) = 1$. Then the map $\psi: M \to M^*$ taking v_j to w_j^*, and w_j to $-v_j^*$ is an isomorphism. The associated product on M, has the property that $\langle v_j,v_i\rangle = 0$, for all i,j, $\langle v_j,w_i\rangle = 0 = \langle w_i,v_j\rangle$ if $i \neq j$, while $\langle v_j,w_j\rangle = 1$ and $\langle w_j,v_j\rangle = -1$. This product is neither symmetric nor positive, but it is non singular. This "symplectic" pairing actually occurs in topology as the intersection pairing of loops on a compact oriented two dimensional surface, such as a "Riemann surface" or smooth complex algebraic curve, of "genus" n. The surface has n holes in it, the v_j represent loops going around the holes in one direction, and the w_j are loops which go around the holes in the other direction.



Ok, back to work

Definition: A **symmetric**, or **self adjoint** operator on a real inner product space M is an endomorphism $T: M \to M$ such that $T = T^*$, i.e. such that $\langle Tx,y\rangle = \langle x,Ty\rangle$ for all x,y in M. On $\mathbb{R}^m$ with the usual

inner product, such an operator is given by a symmetric matrix, i.e. one that equals its transpose.

Definition: An endomorphism $T:M \to M$ of a real inner product space is called **orthogonally diagonalizable** :iff M has an orthonormal basis of eigenvectors of T

Corollary (spectral theorem for symmetric operators): If $T:M \to M$ be a real symmetric endomorphism, then T is orthogonally diagonalizable. [For the converse see Ex. 150.]
proof: The main point is to show that χ_T splits over $\mathbb{R}$. Choose an orthonormal basis for M so that T has a real symmetric matrix A. Then consider A as a complex matrix operating on $\mathbb{C}^m$, by the inclusion $Mat_{m\times m}(\mathbb{R}) \subset Mat_{m\times m}(\mathbb{C})$. Since A is symmetric and real A is hermitian, hence all its characteristic roots are real. Since χ_A is the same whether we consider A as a real or a complex matrix, χ_A splits over $\mathbb{R}$.
Now we can proceed by induction. I.e. if λ is any characteristic root, let v be an associated eigenvector. Then for any w in $v^{\perp}$, we have $0 = \langle v,w\rangle = \langle \lambda v,w\rangle = \langle Tv,w\rangle = \langle v,Tw\rangle$. Hence $v^{\perp}$ is a T-invariant (and thus also T^*-invariant) subspace. The restriction is thus again self adjoint, so by induction has an orthonormal basis of eigenvectors Adding the unit vector $v/|v|$ to this basis, gives an orthonormal eigenbasis for all of M. **QED**.

Definition: An **orthogonal** transformation on a real inner product space M is an endomorphism $T:M \to M$ such that $T^{-1} = T^*$, i.e. such that $\langle Tx,Ty\rangle = \langle x,y\rangle$ for all x,y in M.

Remark: On $\mathbb{R}^m$ with the usual inner product, such an operator is given by an **orthogonal matrix** A, i.e. one whose columns form an orthonormal basis, equivalently whose rows form an orthonormal basis, equivalently such that $A^{-1} = A^t$.

Definition: The set of all orthogonal m×m real matrices forms a group, the **real orthogonal group** O(m). The subgroup of elements of determinant one, is called the **special orthogonal group** SO(m).

Remark: An endomorphism of $\mathbb{R}^m$ with matrix A is orthogonally

diagonalizable iff there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Remark: As observed earlier, a 90^0 rotation matrix about the z-axis in $\mathbb{R}^3$ is an example of an orthogonal matrix which is not diagonalizable over $\mathbb{R}$.

Exercise #161) a) Prove the converse of the previous corollary: if an operator on a finite dimensional real inner product space is orthogonally diagonalizable, then it is self adjoint.
b) For each matrix below, find a matrix, orthogonal if possible, that diagonalizes it over $\mathbb{R}$, or explain why this is not possible:

(i) $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$ (iii) $\begin{bmatrix} 0 & 5 \\ 0 & 0 \end{bmatrix}$ (iv) $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ (v) $\begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$

(vi) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ (vii) $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$ (viii) $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.