Math 4050 (VIGRE) summer 2008 Advanced Linear Algebra

Introduction

These are the slightly expanded classroom lectures from the VIGRE course on "advanced linear algebra" given in Summer semester 2008. It was regarded as a second or upper level course in undergraduate linear algebra, but not a graduate level course. Thus the philosophy was to assume students had experience with the computational aspects of matrices such as row reduction and explicit solving of linear systems, but possibly not with the proofs. Thus the foundations were quickly reviewed with complete proofs of such basic results as the fact that (finite) dimension is well defined as the cardinality of a basis. This theorem received two independent proofs in addition to the fact that it follows from the theory of row reduction.

We gave a complete self contained treatment of the main subject matter of the course, the existence and uniqueness of the Jordan form for operators whose minimal polynomials split over the base field, as well as the generalized Jordan form (a version of the rational canonical form) for linear operators with arbitrary minimal polynomials. In keeping with our philosophy on the level of the course, we used extensively the theory of the Euclidean algorithm for polynomials over a field, but stopped short of employing the language and tools of module theory. Thus we are dealing at all times with elements of the ring k[T] generated by a single operator T over the base field k, and the trick of substituting T for a variable t in a polynomial f(t), but not the abstract module structure over the polynomial ring k[t]. I.e. we use polynomials but not modules or maps of modules, nor exact sequences.

Because of the complexity of the material, we give three proofs of the structure theorems for Jordan and generalized Jordan forms. This allows us to introduce gradually the tools needed in the proofs, and also lets the reader absorb easier independent proofs for the easier cases. I think this is more palatable than trying to understand the hardest cases first and then specialize them to easier situations. This means we treat separately the two extreme cases for Jordan forms, namely when the minimal polynomial f is irreducible, and when the minimal polynomial f equals the characteristic polynomial. In the first "diagonalizable" case the number of cyclic factors in the Jordan decomposition is maximal and each factor is isomorphic to k[t]/f, and in the second case it is minimal, namely there is a cyclic basis and hence only one cyclic factor. The general proofs can be read first, but they seem easier to me if read after these easier warm up cases.

The other standard structure theorems treated, namely the spectral theorems, also receive special proofs for the easer Hermitian/symmetric and unitary/orthogonal cases, as well as later general proofs covering all normal operators, both real and complex. Some discussion of the basic definitions of dual spaces and dual maps are included, as well as a complete self contained treatment of determinants and two proofs of the Cayley Hamilton theorem, one using the structure theorems, and one using only the fundamental formula for expanding a determinant along a column or row and the non commutative factor theorem. Finally we append a list of some good standard and less standard references.

Roy Smith

Table of contents, p.2.

Math 4050 (VIGRE) summer 2008 Vector spaces and linear maps

A vector space is a set of elements which one can add, and also multiply by scalars from some fixed field k, satisfying the usual properties. Precisely, V is a set, +:VxV-->V is a binary operation "addition", k xV-->V is another binary operation "scalar multiplication", and there are 8 properties which must hold: (The reader is encouraged to skip this list of properties and just proceed. She will never miss a thing.)

The first four relate to addition. Addition:
1) is associative, i.e. for all x,y,z in V, (x+y)+z = x+(y+z);
2) and commutative: for all x,y in V, x+y = y+x;
3) has an identity: for some O in V, O+x = x, for all x in V.
4) and has inverses: for each x in V, there is some y in V with x+y = O.

The next four involve scalar multiplication. Multiplication satisfies:
5) "associativity": for all a,b, in k, and all x in V, (a+b)x = ax + bx;
6) "unitary": for all x in V, 1x = x;
7) distributive over addition in k: for all a,b in k, all x in V, (a+b)x = ax + bx;
8) distributive over addition in V: for all x,y in V, all a in k, a(x+y) = ax + ay.

These are similar to the properties of a field. I.e. the first 4 properties above say that (V,+) is an abelian group. Recall a field is a set k with two operations, addition +:kxk-->k, and multiplication kxk-->k, such that (k,+) is an additive abelian group, and (k-{0}, mult.) is a multiplicative abelian group, and multiplication distributes over addition, i.e. x(y+z) = xy + xz, for all x,y,z, in k.

Thus every field k is a vector space over itself. And if k is a field contained in another field L, where the operations in k are the same as the operations in L, then L is a vector space over k. Thus the complex numbers C form a vector space over the reals R, and both of these are vector spaces over the rationals Q.

Exercise: If S is any set, V is a vector space over k, and Fun(S,V) is the set of all functions on S with values in V, define addition of functions pointwise, i.e. (f+g)(p) = f(p) + g(p), for all f,g in Fun and all p in S; and (cf)(p) = c(f(p)), for all c in k, all f in Fun, and all p in S. Then show Fun(S,V) is a vector space with these definitions.
In particular, Fun(S,R) is a vector space over R.

Next we give a way to bootstrap from knowing that the vector space axioms hold for one set (V',+,mult) with 2 operations, to deducing they are also true for another set (V,+,mult) with 2 operations. Assume there is a one one correspondence between V and V' where x in V corresponds to x' in V', and where x+y = z in V, iff x' + y' = z' in V', and cx = y in V if and only if cx' = y' in V'. This called an isomorphism between (V,+, mult) and (V',+, mult), or between V and V' for short.

We claim that if (V',+, mult) satisfies the axioms for a vector space over k, then V does as well. For example, if we want to prove that x+y = y+x in V, it suffices since the correspondence is one to one, to show that (x+y)' = (y+x)' in V'. But (x+y)' = x' + y' = y' + x' = (y+x)'. If we want to show (x+y)+z = x + (y+z), it suffices to show that ((x+y)+z)’ = (x + (y+z))’. But ((x+y)+z)’ = (x+y)’+z’ = (x’+y’)+z’ = x’+(y’+z’) = x’+(y+z)’ = (x+(y+z))’. QED. All the other properties are proved the same way.

Now we can show several different versions of the familiar space of vectors in the plane all satisfy the vector space axioms.

Incarnations of the space of vectors in the Euclidean plane

1) space of plane translations: Let ∏ be the classical Euclidean plane and let T be the set of all translations of ∏, i.e. fixed point free, distance preserving, orientation preserving maps ∏-->∏, with addition defined by composition. Scalar multiplication by integers is defined by repeated addition, multiplication by rationals is by subdivision, and multiplication by reals is achieved by taking limits of rational multiples.

2) equivalence classes of arrows in the plane: T is isomorphic to the set A of equivalence classes of arrows in ∏, where two arrows are equivalent if they have the same direction and length, and addition is defined by adding two representative arrows v,w where the foot of w is at the head of v. Then their sum v+w is the arrow going from the foot of v to the head of w. Scalar multiplication is defined as in T.

3) coordinate vectors in the plane with origin: If we fix an origin O in ∏, then both T and A are isomorphic to the set ∏ where a translation t corresponds to the point t(O), addition is by the parallelogram law as in 2), and multiplication of the point P by r, is by scaling the arrow OP by r. Then T, A and ∏ are all isomorphic.

4) the coordinate plane: If we choose also a unit length in ∏ and a pair of orthogonal axes, then each point P in ∏ has a pair of coordinates (x1,x2), and we get an isomorphism between (∏,O) and R^2, the familiar set of all ordered pairs of real numbers, with component wise addition and scalar multiplication, i.e. (x,y) + (x’,y’) = (x+x’,y+y’), for all real numbers x,y,x’,y’, and c(x,y) = (cx,cy) for all c,x,y in R.

5) functions on a two point set: We claim R^2 is isomorphic to Fun({1,2},R). Namely if x is a function from {1,2}-->R, let it correspond to the pair of values (x(1), x(2)) in R^2; and if (x1,x2) is a pair in R^2, let it correspond to the function x:{1,2}-->R such that x(1) = x1, and x(2) = x2. This a one one correspondence which we easily check defines an isomorphism. E.g. if the functions x,y correspond to the pairs (x1,x2) and (y1,y2), then since addition of functions is by adding their values, the function x+y corresponds to the pair (x1+y1, x2+y2) = (x1, x2) + (y1, y2).

Thus by the previous correspondence principle, since we know Fun({1,2},R) satisfies the vector space axioms over R, so do all the isomorphic spaces 1) - 5) above. Notice that the familiar coordinate plane R^2 in 4) has considerably more structure than the bare Euclidean plane ∏, but that it suffices to choose only an origin as in 3) to view points of the plane ∏ as vectors. The versions 1) and 2) show how to represent vectors in the plane with no arbitrary choices at all. I.e. vectors are not naturally points of the classical Euclidean plane ∏, but are naturally translations of ∏.

Subspaces: If we have a vector space V, and a subset W of V which is closed under the operations on V, we can check that W itself is also a vector space under these operations. Most of the axioms for V guarantee that certain properties hold for all elements of V, hence also for all elements of W. The only two axioms that might fail are those saying that certain elements must exist in W, not just in V. Namely the identity element O must actually be in W. And since W is closed under scalar multiplication, if x is any element in W, then 0x is also in W, and one can show that 0x = O. The other needed thing is that for each x in W, its inverse –x must also be in W. But again W is closed under multiplication so if x is in W, so is (-1)x which one can show equals –x.

Definition: A subset W of a known vector space V, is a subspace of V, if and only if W is itself a vector space with the same operations as in V, i.e. if and only if:
1) O is in W, (or just assume W is non empty), and
2) W is closed under addition and scalar multiplication.

E.g., the space C(R) of continuous R valued functions on R is a subspace of Fun(R,R). For all n > 0, the space C^(n)(R) of R valued functions with n continuous derivatives is a subspace of C(R), as is the space S(R) of infinitely differentiable or "smooth" functions. The space A(R) of analytic functions (defined by convergent Taylor series) is a subspace of S(R). Finally the space P(R) of polynomial functions on R (functions represented by a finite Taylor series) is a subspace of A(R).

Exercise: Check that in a vector space V, 0x = O, and (-1)x = -x. [Hint: 0x = (0+0)x, and 0x = (1-1)x.].] Check also that O is unique and for each x, that –x is unique. [E.g. if O' is another identity, then O = O+O' = O'. Do you see why?]

Linear transformations
Subspaces occur primarily as solution sets to linear problems, i.e. kernels of linear maps. A function or map T:V-->W from one space with addition and scalar multiplication/k, to another, is linear if it takes operations in V into the corresponding operations in W, i.e. if
1) for all x,y in V, T(x+y) = T(x) + T(y),
2) for all c in k, and all x in V, we have T(cx) = cT(x).

It is often helpful to study maps as compositions of simpler maps.
Define: The composition of two linear maps S:V-->W, and T:W-->U, denoted ToS:V-->U, or just TS:V-->U, (if this will not be confused with some other type of multiplication), is defined as usual by (TS)(x) = T(S(x)).

Ex: Check that TS is linear if both S,T are linear.

Definition: In particular, T:V-->W is an isomorphism between V and W, if there is also a linear map S:W-->V which is inverse to T:V-->W, i.e. ST = Id on V, and TS = Id on W, where ST denotes composition, i.e. ST(x) = S(T(x)).

Recall: A map T:V-->W is injective if for every y in W there is at most one x in V with T(x) = y. T:V-->W is surjective if for every y in W, there is at least one x in V with T(x) = y. T is bijective if it is both injective and surjective.

Ex: A linear map T:V-->W is bijective if and only if it has a linear inverse, i.e. is an isomorphism. [Hint: Show the inverse function of a bijective linear map, is also linear.]

Normally we assume when given a linear map T:V-->W, that both V and W are vector spaces, i.e. both satisfy the 8 axioms, but in fact we can generalize the correspondence principle at least to the case of a surjective linear map, to prove this.

Exercise: If T:V-->W is a surjective linear map between spaces with two operations as above, and if V satisfies the vector space axioms, then so does W.
[E.g. to show W has an additive identity, consider the element T(O). Then for every y in W, there is some x in V with T(x) = y. Then y = T(x) = T(x+O) = T(x) + T(O) = y + T(O). Hence T(O) acts as an additive identity in W.]

Many linear maps are not isomorphisms, e.g. they may not be injective. The following concept measures the possible failure of injectivity.
Definition: If T:V-->W is a linear map of vector spaces, define kernel(T) = ker(T) = N(T) = nullspace(T) = {the set of x in V such that T(x) = O}.

Lemma: ker(T) is a subspace of V.
proof: Since O+O = O, T(O) = T(O+O) = T(O) + T(O), so by subtracting T(O) from both sides, we see O = T(O). You check ker(T) is closed under both operations. **QED.**

Example: Let C^inf(R) = S(R) for short, denote the space of infinitely differentiable, or "smooth", functions R-->R, and define D:S(R)-->S(R) to be differentiation, i.e. Df = f'. The mean value theorem of calculus implies ker(D) = R = constant functions on R.

Example: If a is a real number define (D-a):S(R)-->S(R) by (D-a)f = f'-af. Then ker(D-a) = {smooth f: f' = af}. Recall a constant multiple f = ce^(at), of the exponential function e^(at) has this property. We claim these are the only solutions. Proof: If f' = af, then the derivative by the quotient rule of f/e^(at) is [f'e^(at) – fae^(at)]/e^(2at) = [afe^(at) – fae^(at)]/e^(2at) = 0, so by MVT, f/e^(at) = c is constant, i.e. f = ce^(at).

Lemma: If T:V-->W is a linear map, then T is injective if and only if ker(T) = {O}.
proof: If T is injective then since we already know T(O) = O, then T(x) cannot equal O for any other x, so ker(T) = {O}. Conversely, if ker(T) = {O}, and Tx = Ty, then T(x-y) = O, so x-y is in ker(T) = {O}, so x-y = O, and x= y. Hence T is injective. **QED.**

Cor: If T:V-->W is surjective and linear, T is an isomorphism iff ker(T) = {O}.
proof: Then ker(T) = {O} implies T is bijective and linear. **QED.**

Image spaces
We can also measure the failure of surjectivity, by a subspace of the target. If T:V-->W is linear, put Image(T) = Im(T) = Range(T) = R(T)= {those y in W such that there is at least one x in V with T(x) = y}, the set of values actually taken on by T in W.

Lemma: For every linear map T:V-->W the subset Im(T) is a subspace of W.
proof: In fact since by definition of Im(T) the map T:V-->Im(T) is surjective, it follows as in the exercise above that Im(T) is closed under both operations on W, and satisfies the vector space axioms since V does. **QED.**

Clearly T:V-->W is surjective iff Im(T) = W. Surjectivity is the abstract version of solvability of a problem. I.e. giving y is the problem, and x with Tx = y, is a solution. Soon we will learn to measure the "dimension" of a subspace, and then we will often be able to tell that Im(T) = W, i.e. all problems have solutions, by computing dimensions.

Understanding linear isomorphisms k^n-->V, equivalently "bases" of V
We will prove that a "finite dimensional" vector space V over k is isomorphic to a unique coordinate space k^n, which will let us make calculations in arbitrary spaces using numbers. In particular the integer n will measure "how big" the space V is, i.e. its "dimension". Thus we want to learn to define linear maps k^n-->V and tell when the map is an isomorphism. This leads to the concepts of independence and span, which correspond to injectivity and surjectivity.

Product spaces, projections, inclusions, and maps of products
If V,W are two vector spaces over k, define operations componentwise on their cartesian product set VxW = {all ordered pairs (x,y): where x in V, y in W}, i.e. (x1,y1) + (x2,y2) = (x1+x2, y1+y2), and c(x,y) = (cx,cy). E.g. R^2 = RxR.

Exercise: VxW is a vector space over k, if V,W are.

Example: We define products of more factors similarly, e.g. k^n = kx....xk (n factors) = {(x1,...xn): all xi elements of the field k}, is the usual Euclidean coordinate n - space.

Define: The projection π2:VxW-->W by π(x,y) = y. Then π2 is surjective, linear and ker(π2) = {(x,0): all x in V} = Vx{0}, which is isomorphic to V by the obvious correspondence x <---> (x,0). Similarly π1:VxW-->V is defined by π1(x,y) = x.

Define: Inclusions j1:V-->VxW by j1(x) = (x,O), and j2:W-->VxW by j2(y) = (O,y). These are injective, linear and Im(j1) = Vx{O} isom to V, Im(j2) = {O}xW isom to W.

The key to defining maps on a product is the following:
Theorem: If V,W,U are vector spaces /k, then h:VxW-->U is a linear map iff h is the sum of a pair of linear maps f:V-->U and g:W-->U. The maps f,g are unique
proof: Given h:VxW-->U, if we want h((x,y)) = f(x)+g(y), we must have h((x,O)) = f(x) + g(O) = f(x), and similarly must have h((O,y)) = g(y), so there is only one possible choice for f,g. With this choice, we have f(x) = h o j1, and g(y) = h o j2, so f,g are

compositions of linear maps, hence linear, and h((x,y)) = h((x,O) + (O,y)) = f(x) + g(y). Linearity of a sum f+g is easy. **QED**.

Hence a linear map out of a product may be defined arbitrarily on the factors. So if we understand maps on V and on W, we also understand maps on VxW. The result also holds for more factors. This lets us describe all linear maps k^n-->U:

Define: ei = (0,...,1,....0), where the 1 appears only in the ith position.

Theorem: i) A linear map f:k-->U is defined by f(t) = ty, where y = f(1) is arbitrary. Thus a linear map k-->U corresponds to the choice of y in U, i.e. Hom(k,U) = U.
ii) A linear map f:k^n-->U has the form f(t1,...,tn) = t1 y1+...tn yn, where yi = f(ei). Thus a linear map k^n-->U corresponds to a sequence (y1,...,yn), i.e. Hom(k^n,U) = U^n.
proof: i) If f is linear, then indeed f(t) = f(t1) = tf(1). And if y is arbitrary, defining f(t) = ty, satisfies f(s+t) = (s+t)y = sy + ty, and f(st) = sty = s(ty) = sf(t), is linear. then ii) follows from our previous result about maps out of products. **QED.**

Moral: Thus to define a linear map f:k^n-->U, just choose any sequence (y1,...,yn) of vectors in U, and there is a unique linear map f with f(ei) = yi, for all i=1,..,n.

We want to know when such an f is an isomorphism.
Define: Given a (possibly infinite) sequence (y1,...,yn,......), a sum of scalar multiples of finitely many of these vectors, i.e. an expression of form a1 y1+....an yn, where all ai are in k, is called a "linear combination:" of the vectors in the sequence.

Reminder: Linear combinations are by definition ALWAYS finite.

Theorem: If f:k^n-->U is linear, with defining sequence (y1,...,yn) where yi = f(ei), then f is surjective iff every vector in U equals some linear combination of (y1,...,yn).
proof: If f is surjective, and z in U, then z = f(a1,...,an) for some (ai). But (a1,...,an) = a1e1 + ...+ an en, so z = f(a1e1 + ...+ an en) = a1f(e1) + ...+an f(en) = a1 y1+....an yn. I.e. z is a linear combination of (y1,..,yn). Conversely, if every z in U is such a linear combination, then for each z, we have z = a1 y1+....an yn. Then f(a1 e1+....an en) = a1f(e1) + ...+an f(en) = a1 y1+....an yn = z. So z = f(x) where x = a1 e1+....an en, hence f is surjective. **QED.**

Define: A sequence (y1,...,yn) in U), is said to "span" U, or to "generate" U, if every non zero vector in U has at least one expression as a linear combination of (y1,...,yn).
Thus f:k^n-->U is surjective iff (f(e1),...,f(en)) spans U. We will also say the empty set spans the zero space {O}.

Next we consider the number of ways to express a vector as a linear combination of a given sequence of vectors.
Define: A sequence (y1,...,yn) in U is called "linearly independent" or "independent", if each vector in U has at most one expression as a linear combination of (y1,...,yn). E.g. the empty set is independent.

Theorem: If f:k^n-->U is linear, and yi = f(ei), then f is injective iff the sequence (y1,...,yn) is independent, iff O can be expressed as O = a1y1+...an yn, only in the "trivial way", i.e. iff O = a1y1+...an yn holds only when all ai = 0.
Proof: By definition of injective, since f(a1,..,an) = a1 y1+....an yn, no vector in U can have more than one such expression if f is injective. Moreover we know f is injective iff ker(f) = {O}, hence iff O has only one such expression. Since indeed O = 0 a1+...0 an, f is injective iff this is the only way to represent O as a linear combination of (y1,...,yn).
QED.

Next we want to describe vector spaces which are isomorphic to k^n.
Lemma: If (y1,...,yn) is a sequence which spans U, then there is a subsequence which not only spans U but is also independent.
proof: If (y1,...,yn) is not independent, then some expression a1y1+...an yn = O, where not all ai = 0. After renumbering, say an ≠ 0. Then solving for yn, we have yn = -(a1/an) y1 - ...-(an-1/an) yn-1. Then we claim (y1,...,yn-1) still spans U. I.e. if z is a vector in U written as z = c1y1+...cn yn, to write it in terms only of (y,...yn-1), just substitute for yn, using the equation above yn = -(a1/an) y1 - ... -(an-1/an) yn-1.

If the subsequence (y1,...yn-1) is independent, then we are done. If not, eliminate another vector in the same way. Since at every stage the new smaller subsequence still spans, eventually either we get an independent subsequence that spans, as desired, or we get down to the empty set, which still spans, hence U = {O}, and we are also done. **QED**

Theorem: V is isomorphic to some k^n iff some finite sequence in V spans V.
proof: If f:k^n-->V is an isomorphism, then f is surjective so f(e1),...,f(en) spans V. Conversely if (y1,...,yn) spans U, and if (y1,...,ym) is a subsequence which spans V and is also independent, then the unique linear map f:k^m-->V with f(ei) = yi, is bijective. **QED**

Define: An independent sequence (y1,...yn) in V which spans V, is called a **basis** for V.

Cor: If V has a finite spanning set, then V also has a basis, and f:k^n-->V is an isomorphism iff (y1,...yn) = (f(e1),...,f(en)) is a basis of V.

Define: V is called "finite dimensional" iff V has a finite spanning set, iff V has a finite basis, iff there is an isomorphism k^n-->V for some n ≥ 0.

Next we need to know that V cannot be isomorphic to two different spaces k^n with different values of n. Since composition of isomorphisms is an isomorphism, if V were isomorphic to two different coordinate spaces, those spaces would be isomorphic to each other, so it suffices to prove the following.

Theorem: If f:k^n-->k^m is injective, then n ≤ m, (hence f isomorphic implies n = m).
proof: Induction on m. If m = 1, and n ≥ 2, and f is injective, then f(e1) = s and f(e2) = t, must be independent in k. But any two elements of k are dependent. [If either s or t is zero, then (s,t) is dependent, and if say s ≠ 0, then (t/s)s – t = 0.]

Now assume m ≥ 2, and f:k^n-->k^m is injective. If Im(f) does not contain the vector (0,...,0,1) = em, then composing f with projection to k^(n-1), is still injective k^n-->k^(m-1), hence by induction n ≤ m-1 < m, and we are done.

If f(x1,...,xn) = (0,...,0,1) = em, for some x = (x1,...,xn), then some xi ≠ 0, say xn ≠ 0. Then since f is injective, the only vectors mapped to the subspace spanned by em, are multiples of x. Then for the restriction g of f to k^(n-1), g:k^(n-1)-->k^m, g is still injective, and no longer does Im(g) contain the vector em. Hence composing with projection to k^(m-1) gives an injection k^(n-1)-->k^(m-1) and again by induction we have n-1 ≤ m-1, so n ≤ m. **QED**

Cor: If V is a finite dimensional vector space/k, V is isomorphic to k^n for precisely one integer n ≥ 0, moreover every basis of V has the same number of elements, namely n.

I liked making up that proof but everyone should know Riemann's brilliant argument, called the (Steinitz) replacement lemma, as usual with historical miscrediting of ideas.
Exercise: i) A sequence (w1,...,wt) is dependent iff some wj is a linear combination of vector to the left of it (0 is a linear combination of the empty set).
ii) Any vector in a sequence which is a linear combination of other vectors can be removed without changing the span of the sequence.

Proposition: Given a sequence (x1,...,xn; y1,...,yr; z1,...,zs) in a vector space V assume (x1,...,xn; y1,...,yr) is independent, and (y1,...,yr; z1,...,zs) spans. Then s ≥ n.
proof: Assume n,r,s, ≥ 0 and use induction on n. We are fine if n = 0, so assume n ≥ 1, and that the result holds for n-1 and all r,s. Then just move the last x into the sequence of y's. I.e. consider the sequence (x1,...,xn-1; xn,y1,.....,yr; z1,...,zs). Since by hypothesis (y1,.....,yr; z1,...,zs) spans, the larger sequence (xn,y1,.....,yr; z1,...,zs) both spans and is dependent. Hence by the previous exercise i) some vector in this longer sequence depends on previous vectors. But since the x's and y's are independent, then s ≥ 1 and there exists some z which is dependent on earlier vectors. If we remove that z and renumber we have now (xn,y1,.....,yr; z1,...,zs-1) which still spans, by exercise ii). Then we can apply the induction hypothesis to (x1,...,xn-1; xn,y1,.....,yr; z1,...,zs-1). I.e. now (x1,...,xn-1; xn,y1,.....,yr) is independent, and (xn,y1,.....,yr; z1,...,zs-1) spans, so by induction s-1 ≥ n-1, hence s ≥ n. **QED.**

Cor: If a vector space V has a spanning set with m elements, then no independent set can have more than m elements.
proof: This is the case of the proposition with r = 0. **QED.**

Cor: Any two finite bases for a space V have the same number of elements.

Define: If V is finite dimensional /k, the number of elements in any hence every basis for V, is called the dimension of V.

Cor: If W is a subspace of V, and V is finite dimensional, then W is also finite dimensional, and dim(W) ≤ dim(V), and W = V iff dim(W) = dim(V).

Dimension theory lets us prove surjectivity of maps by computing dimensions.
Cor: If f:V-->W is linear, ker(f) = {O} and dim(W) ≤ dim(V), then f is an isomorphism.
proof: A basis for V goes to an independent set in W with ≥ dim(W) elements, hence a basis for V. **QED**

We can generalize this as follows.
Theorem: If f:V-->W is linear, and dim(W) + dimker(f) ≤ dim(V), then f is surjective.
Proof: Extend a basis for ker(f) to a basis for V, which complementary set of vectors maps to an independent set in W with ≥ dim(W) elements, hence to a basis, so the full basis maps to a spanning set. **QED.**

The proof shows the following.
Cor: For any linear map f:V-->W with dimV finite, dim(ker(f)) + dim(Im(f)) = dimV.

This is sometimes called the **rank - nullity theorem** since if rank(f) = dim(Im(f)), and nullity(f) = dim(ker(f)), then for any linear map f on a finite dimensional space V, **rank(f) + nullity(f) = dim(V)**.

Exercise: If V,W have the same finite dimension, and T:V-->W is linear, then T is injective iff T is surjective iff T is bijective iff kerT = {0}.

Exercise: In a finite dimensional space every spanning set contains a basis, and every independent set is contained in a basis.

Remark: Bases always exist in all vector spaces, even infinite dimensional ones, and any two always have the same possibly infinite cardinality. These proofs are a little tedious but one aspect is easy.

Exercise: If B is a maximal independent set in a vector space V, possibly infinite, i.e. if adding any vector B makes it dependent, then B is a basis for V.

An easy application of "Zorn's lemma" shows maximal independent sets, hence bases, always exist, and that an independent set is contained in a basis, and a spanning set contains a basis. It is harder to show two bases always have the same cardinality.

Matrix computations: composition of maps corresponds to matrix multiplication

Now that we know the theory of dimensions of subspaces, we want to calculate them in practice. E.g. given a linear map we want to calculate the dimensions of its kernel and image, and to find bases for these spaces. There is an elementary way to do this. If bases of V and W are given we represent f by a matrix of elements of k, and then "row reduction" of the matrix computes bases of the subspaces ker(f) and Im(f).

First we discuss maps f:k^n-->k^m, where the bases are the standard ones, (e1,...,en), and (e1,..,em). We represent vectors in k^n by columns of scalars of length n, and vectors in k^m by columns of length m. Since a linear map f:k^n-->k^m is determined uniquely by the ordered sequence (f(e1),...,f(en)) in k^m, we represent f by the corresponding sequence of n columns of vectors, each of length m. Thus the jth column of the matrix is [f(ej)]. Since this rectangular array of scalars has m rows and n columns, we call it an "m by n" matrix (of elements of k). E.g. if f:k^2-->k^3, and f(e1) = 2 e2+ e3, f(e2) = e1 – e2, the matrix of f is this 3 by 2 matrix :

$$\begin{vmatrix} 0 & 1 \\ 2 & -1 \\ 1 & 0 \end{vmatrix}.$$

It is obvious these column vectors f(e1), f(e2) are independent (the first entry is zero in one but not the other) hence f is injective, dimker(f) = 0, and dimIm(f) = 2.

If f has matrix : $\begin{vmatrix} a & x \\ b & y \\ c & z \end{vmatrix}$, i.e. if f(e1) = ae1 + be2 + ce3, and f(e2) = xe1 + ye2 + ze3,

then we know how to compute f(t1 e1 + t2 e2) for any other vector (t1,t2) in k^2. Namely f(t1 e1 + t2 e2) is the linear combination t1 f(e1) + t2 f(e2), which as column vectors is:

$$t1 \begin{vmatrix} a \\ b \\ c \end{vmatrix} + t2 \begin{vmatrix} x \\ y \\ z \end{vmatrix}$$

We define this as a type of multiplication of matrices: i.e. f(t1e1 +t2e2) =

$$\begin{vmatrix} a & x \\ b & y \\ c & z \end{vmatrix} \begin{vmatrix} t1 \\ t2 \end{vmatrix}$$

= the linear combination of the columns of the left hand matrix, with coefficients being the entries of the right hand column matrix.

For this to make sense, the length of the right hand column has to equal the width of the left hand matrix, i.e. the number of rows of the right hand matrix has to equal the number of columns of the left hand matrix. As long as this is true, we can extend it to a right hand matrix with more columns, as long as they all have the same number of rows.

$$\begin{vmatrix} a & x \\ b & y \\ c & z \end{vmatrix} \begin{vmatrix} t1 & s1 & r1 \\ t2 & s2 & r2 \end{vmatrix}$$

First we multiply the first column on the right by the left hand matrix getting a column of length 3 as answer. then we multiply the second column on the right by the left hand matrix, getting a second column of length three. Then we do it again. Thus the result is a matrix having three columns each of length 3. The first column in the product is the linear combination of the left hand columns with coefficients from the first column on the right, namely this column vector:

$$t1 \begin{vmatrix} a \\ b \\ c \end{vmatrix} + t2 \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} at1+xt2 \\ bt1+yt2 \\ ct1+zt2 \end{vmatrix}$$

The second column in the matrix product is the linear combination of the left hand columns with coefficients from the second column on the right, and so on. So the product matrix looks as follows:

$$\left|\begin{array}{cc} a & x \\ b & y \\ c & z \end{array}\right| \left|\begin{array}{ccc} t1 & s1 & r1 \\ t2 & s2 & r2 \end{array}\right| =$$

$$\left|\begin{array}{ccc} at1+xt2 & as1+xs2 & ar1+xr2 \\ bt1+yt2 & bs1+ys2 & br1+yr2 \\ ct1+zt2 & cs1+zs2 & cr1+zr2 \end{array}\right| .$$

There are three columns because they arise from the three columns of the right hand matrix, and three rows because the matrix on the left had three rows. The three columns are the result of applying the map represented by the left hand matrix, consecutively to the three column vectors in the right hand matrix.

A simple mechanical way to look at this is as follows: the jth column in the product is a linear combination of the columns of the left factor, with coefficients from the jth column of the right factor. Since the ith entry of a linear combination of vectors is just the same linear combination of their ith entries, it follows that the ith entry of the jth column of the product is a linear combination of the elements of the ith row of the left factor with coefficients from the jth column of the right factor. Notice in particular, the rows of the left factor are the same length as the columns of the right factor. This gives the following rule:

Define the dot product of two vectors in k^m as follows:
(a1,....am).(t1,...t,m) = a1t1+a2t2+....+amtm. Then two matrices A,B can be multiplied in the order AB, if the rows of A are the same length as the columns of B, and then the entry in the ith row and jth column of the product AB, is simply the dot product of the ith row of A, with the jth column of B.

E.g. $$\left|\begin{array}{ccc} 2 & 3 & -1 \\ 3 & 4 & 5 \end{array}\right| \left|\begin{array}{ccc} 1 & 0 & 2 \\ 7 & 1 & 9 \\ -2 & 8 & 6 \end{array}\right| = \left|\begin{array}{ccc} 27 & -5 & 25 \\ 21 & 44 & 72 \end{array}\right|.$$

We looked at matrix multiplication AB as forming linear combinations of the columns of A, but the final rule for multiplying with dot products also used rows. Thus the process has another aspect. I.e. suppose we think of forming a new matrix AB by telling how to form its rows. Namely we want to the rows to be linear combinations of the rows of B, with coefficients from the rows of A. I.e. the ith row of AB is the linear combination of all rows of B, using as coefficients the entries in the ith row of A. Then what is the jth entry of that row? Well it is obtained by forming the linear combination of the jth entries of all the rows of B, with the same coefficients, namely the entries from the ith row of A. But this is again the dot product of the ith row of A with the jth column of B. Thus the two ways of defining a matrix product AB, by forming linear combinations of the rows of B, or linear combinations of the coumns of A, give the same result, and we can regard the process either way that is helpful.

The link with the concrete problems of understanding ker(T) and Im(T) is this: If A is the matrix of T:k^n-->k^m, i.e. if the column vectors of A are (T(e1),...,T(en)), then Im(f) = Span(f(e1),...,f(en)) = Span of the columns of A, called the "column space" of A. This is a subspace of k^m. On the other hand, Ker(T) = the space of all vectors v in k^n such that Av = O. Here v is a column vector in k^n.

Concretely, if A = $\begin{vmatrix} 2 & 3 & -1 \\ 3 & 4 & 5 \end{vmatrix}$ is the matrix of T:k^3-->k^2, then Im(T) is the subspace of k^2 spanned by (2,3), (3,4), (-1,5), i.e. all of k^2. Ker(T) is the space of solutions (x,y,z) of the simultaneous system of equations 2x + 3y –z = 0 = 3x +4y +5z. Notice the equations are given by the rows of A. Since we know the kernel is one dimensional, it suffices to find one non zero solution of these equations. Multiplying the first equation by 3 and adding that to minus twice the second equation, leaves the equations: 6x + 9y - 3z = y -13z = 0. Now x has been eliminated from the second equation, so we can set z = 1, then the second equation is satisfied by y = 13, and the first one then is also satisfied by taking 6x = 3 -9(13) = -114, hence x = 114/6 = 19.

The fundamental fact: matrix multiplication corresponds to map composition

The product AB of two matrices, represents the composition of the maps represented by the matrices A and B. To see this assume A represents the map S, and B represents the map T. Now recall that the matrix of the composition ST has as columns the images of the standard basis vectors ej under the map ST. But the columns of B are the images of the ej under the map T, and the columns of AB are images of the columns of B under the map S, i.e. the columns of AB are indeed the images of the ej under the composition ST. In particular, since composition of maps is associative, so is matrix multiplication. Thus the set Mat(nxn;k) of n by n matrices over a field k forms a ring which contains k via the map c-->cId.

Cor: If L(V) is the set of k linear maps V-->V, a choice of basis (x1,...,xn) for V defines a ring isomorphism L(V)-->Mat(nxn;k).

proof: If Q:k^n-->V is the isomorphism sending ej to xj, then sending T in L(V) to the matrix for Q^-1TQ is the corresponding isomorphism. **QED**

Summary of theory of row reduction

We assume the reader knows the elementary theory of row reduction for matrices. This consists in using three elementary row operations to reduce a matrix to echelon form, which has the same kernel as the original matrix. The three operations are:

i) interchange any two rows,

ii) multiply any row by a non zero scalar,

iii) add any multiple of one row to a different row.

These operations do not change the kernel of the matrix and suffice to reduce the matrix to one such that all zero rows are at the bottom, and in every non zero row the first (left most) non zero entry occurs to the right of the first non zero entry in the previous row. Such a matrix is said to be in echelon form, and a basis for its kernel is easily given, starting at the bottom and solving upwards. We call a column a pivot column if it contains the first non zero entry of some row. Then one can assign arbitrary values to entries corresponding to non pivot columns, and the entries corresponding to pivot columns are uniquely determined. It follows that the kernel is isomorphic to k^d where d = number of non pivot columns.

Each non pivot column can be expressed in this way as a linear combination of the previous pivot columns, the ones to its left, by findinf an element of the kernel with entry -1 at the given non pivot column and zeroes at the others and all later columns. This allows us to explicitly reduce a sequence of vectors to an independent sequence with the same span. Namely one puts the sequence in as columns, reduces the matrix to echelon form, and then goes back to the original matrix to select those columns in the same position as the pivot columns from the reduced matrix. Since reduction explicitly reveals a non zero element in the kernel of any matrix having more columns than rows, it also proves our basic dimension result that no linear map from k^n-->k^m can be injective if n > m. The location of the pivot columns is also determined independently of the manner of reduction since a column is a pivot if and only if it does not depend linearly on previous columns.

Row reduction also allows one to calculate the inverse of an invertible n by n matrix. Just augment that matrix by putting a copy of the n by n identity matrix to its right, and reduce the whole n by 2n matrix to reduced echelon form (where each pivot column has only one non zero entry, a '1' in the pivot position). Then the original matrix A, if invertible, has become the identity, and the identity matrix to its right has become A^-1. This happens because row operations are equivalent to left multiplication by certain elementary matrices, hence reduction to the identity must be multiplication by the inverse matrix. To find the elementary matrix which achieves a certain row operation, just perform the operation on the identity matrix, and then left multiplying by the resulting matrix will perform that row operation. Since left multiplication by an invertible matrix corresponds to composing with an injective map, it does not annihilate any non zero vector, hence row operations do not change the kernel of a matrix.

Personal remark: As a student I learned mostly Banach spaces and infinite dimensional spectral theory, not the uses of row reduction, and was amazed as a young teacher to learn that in finite dimensions one can explicitly solve linear equations! The other professors and I had fun educating each other. They taught me trigonometry, calculus, linear algebra, and Galois theory, and I taught them differentiable manifolds, categories, functors, tensor products, sheaves and deRham cohomology. We were puzzled at each other's ignorance. I couldn't understand why they had PhD's and I didn't, so I went back to grad school. Then I found out it's not what you "know", but what you can do. They had all found new results and proved them, and I hadn't. Moral: whatever you study, learn to understand it, how to use it, make calculations, and look for open questions.

Matrices for linear maps of finite dimensional vector spaces

If T:V-->W is any linear map of finite dimensional spaces, we want to represent T as a matrix. To do this we choose bases in both V and W, hence isomorphisms P:k^n-->V and Q:k^m-->W, and by composition a linear map (Q^-1oToP):k^n-->k^m. By our previous discussion, this composition has a matrix, called the matrix of T with respect to the given bases. In particular if T:k^n-->k^m is originally given by a matrix A, choosing new bases in k^n and k^m gives a new matrix Q^-1 A P for A. Our goal is to determine bases which make the matrix of T as simple as possible.

Note: The matrix of T:V-->W in the bases (x1,...,xn) for V and (y1,...,ym) for W, has jth column equal to the coefficients of the vector T(xj) in terms of the basis (y1,...,ym).

We examine some important examples next.

Matrices for projections, reflections, stretches, and shifts

Theorem: A linear map T:V-->W between finite dimensional spaces has a diagonal matrix for some choices of bases in V and W, with only 1's and 0's on the diagonal.

proof: Assume V has dimension n, ImT has dimension k, hence kerT has dimension n-k by the rank nullity theorem; let (xk+1,...,xn) be a basis for kert, and extend it to a basis (x1,...,xk,xk+1,...,xn) for V. Then we know (T(x1),...,T(xk), T(xk+1),...,T(xn)) spans ImT. But since the last n-k of these vectors are all zero, [i.e. (xk+1,...,xn) is a basis for kerT, so T(xk+1) = ... = T(xn) = 0], it follows that the first k of them, namely (T(x1),...,T(xk)), already span ImT. Since we also know ImT has dimension k, these vectors T(x1) = y1, ..., T(xk) = yk, form a basis for ImT, which we can then extend to a basis (y1,...yk,yk+1,...,ym) for W. Since then T(xi) = yi for 1 ≤ i ≤ k, and T(xj) = 0 for j = k+1,...,n, the matrix of T in these bases is as described. I.e. it has a k by k identity matrix in the upper left corner, and all the rest of the entries are zeroes. **QED.**

For example, if k = 2, n = m = 4, we have: $$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since that matrix is as simple as possible, and contains only the information of the rank and nullity of the original map, we ask what can be said about the matrix of a map T:V-->V from a finite dimensional space to itself, in terms of one basis (x1,...,xn) for V; how simple can it be made, and what information will it contain about T? This means we restrict ourselves to using the same basis isomorphism Q:k^n-->V at both ends of the composition discussed above. So given T:V-->V, we seek an isomorphism Q:k^n-->V such that the composition Q^-1TQ has the simplest possible matrix. Or if A is already an n by n matrix, we seek an invertible n by n matrix Q such that Q^-1AQ is as simple as possible.

For example, which maps V-->V have a matrix as simple as the one just above? Looking at it, and recalling that the first column is the expansion of f(x1) as 1x1 + 0x2 + 0x3 +...., we see that T must take the first basis vector itself, and the same for the second through the kth vector. Then T must send the last n-k vectors to zero. So we have a basis (x1,...,xk,xk+1,...,xn) such that T(xi) = xi, for 1 ≤ i ≤ k, and T(xj) = 0, for k+1 ≤ j ≤ n.

Then if V1 = Span(x1,...,xk), and V2 = Span(xk+1,...,xn), we have V isomorphic to V1xV2, and T corresponds to the identity on V1, while T corresponds to 0 on V2. I.e.

T corresponds to the projection V1+V2-->V1, taking (x,y) to x. We call T a projection of V onto the subspace V1 "along" the subspace V2. Notice then that V1 = ImT, and V2 = kerT. So for T to be a projection, T must equal the identity on ImT, which implies of course that kerT and ImT intersect only in {0}, (since if T(x) = x and T(x) = 0, then x = 0). Hence we call this matrix a projection matrix. We have essentially proved:

Theorem: T:V-->V has in some basis for V a diagonal matrix with 0's and 1's on the diagonal, iff T = Id on Imf, iff T^2 = T, iff T is a projection onto ImfT along kerT.

Proof: Note that if T = Id on ImT then for all x in V, T(x) is in ImT so T(T(x)) = T(x), i.e. T^2 = T, and conversely if T^2 = T, then for all x in V, T(T(x)) = T(x), so T = Id on ImT. **QED.**

Remark: Thus if T:V-->V is linear, then there is a basis for V in which T has a projection matrix as above if and only if T^2 = T. Notice then that T^2 –T = 0, i.e. T satisfies the polynomial t^2 - t = 0. In general, finding the appropriate polynomial satisfied by a matrix will tell us much about the structure of the correspon map.

What about a diagonal matrix with no 0's on the diagonal, but only 1's and -1's? E.g. when can the matrix for T look like $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$?

This would say that T fixes the vectors x1 and x2, and sends x3 to –x3. So T "reflects" everything in the plane spanned by (x1,x2), except the reflection is not necessarily a perpendicular one, but reflection "along" the axis spanned by x3. We claim this occurs (except in characteristic 2) if and only if the dimensions of the kernels of (T-1) and (T+1) add up to dimV.

Note: If T is a reflection, then T satisfies the polynomial (t-1)(t+1) = 0, (but not any proper factor of it). We will see the converse statement holds as well.

Shift (nilpotent) operator: It is easy to see the matrix of the derivative operator D acting on the space of polynomials of degree ≤ 3, with basis {t^k/k!}, 0 ≤ k ≤ 3, looks as follows: $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. When does this happen? T has such a matrix, if and only if there is a basis (x1,...,xn) for V such that f(xi) = xi-1, for 2 ≤ i ≤ n, and f(x1) = 0.

Remark: Such a T satisfies the polynomial t^n = 0, where n = dimV, but satisfies no proper factor, and we will see this fact characterizes a shift operator.

Stretches (diagonalizable operators):
A linear map T:V-->V is a “stretch” if there is a basis (x1,..,xn) for V such that for all i, T(xi) = ci xi, is a scalar multiple, or stretch, of xi, (of course ci could be negative). In such a basis, the matrix of T is diagonal, with the stretching factors ci on the diagonal.

Such a map satisfies the polynomial ∏(T-ci) = 0, where only distinct ci appear. E.g. On the space spanned by (e^c1t,...,e^cn t) the derivative operator D has diagonal matrix with (c1,...,cn) on the diagonal. Here is such a diagonal matrix representing the derivative D on the space Span(e^t, e^2t, e^-t), and satisfying (t-1)(t-2)(t+1) = 0.
| 1 0 0 |
| 0 2 0 |
| 0 0 -1|.

Stretches plus shifts
In many settings the most general operator can be represented as a sum of the last two examples, as a stretch plus a shift. We will study examples where D has such a matrix more closely in the next paragraph. E.g. on Span(e^2t, te^2t, (1/2)t^2e^2t) the matrix for D looks as follows:
| 2 1 0 |
| 0 2 1 |
| 0 0 2 |.

We will prove every operator satisfying a polynomial which factors over k into linear factors has a matrix representation composed of blocks of stretches, shifts, or their sums, and if the factors are distinct, it has a diagonal matrix. This is called the Jordan form. For brevity we call all such matrices stretch plus shift.

Solution spaces of linear differential operators with constant coefficients.
We study next differential equations of form: an y^(n) +...+ a1 y' + a0 y = f(t) where f is a given polynomial, and y is a smooth function to be found. One reason for studying this problem, aside from its intrinsic importance, is the insight it will give us into the structure of linear operators on finite dimensional spaces. It turns out that almost every operator on a finite dimensional space, looks like copies of the operator D on the solution space of a differential equation like the one above.

Recall we define the operator D on the space S(R) of "smooth" functions f:R-->R, by setting D(f) = f', the derivative of f, and then D is linear, hence an element of the space Hom(S(R)) = L(S(R)) = {all linear operators S(R)-->S(R)}. Notice this space of operators is a vector space, but it also has another operation defined, composition. Composition is a kind of multiplication that distributes over addition since the operators are linear, and this makes L(S(R)) into a ring containing the constants R, which act by multiplication. Thus L(S(R)) is an "algebra" over R, (i.e. ring containing R). In this algebra we consider the subalgebra R[D] generated by D and the constants. This consists of all operators of form: an D^n +...+ a1 D + a0, where the ai are constants. Since this is just a polynomial in D with real coefficients, we may denote it as P(D), where P(X) is the polynomial an X^n +...+ a1 X + a0. In particular the differential equation above:
an y^(n) +...+ a1 y' + a0 y = f(t), may be written as P(D)y = f. So we are interested in solving linear equations of this form P(D)y = f , where P(D) is a linear operator in the subalgebra R[D], of linear operators on S(R) generated by D, and f is a polynomial.

The following lemma is fundamental:

Lemma: If T:V-->W is any linear map, and v a vector in V such that Tv = w, then the full set of solutions x of the equation Tx = w, consists of those vectors x such that x-v belongs to kerT. I.e. Tx = w if and only if x =v +u where u is in kerT.
proof: An easy, but important, exercise.

Thus an equation of form Tx = w has either no solutions, or has as many as there are elements of kerT. Then the problem of solving Tx = w, has two parts:
i) finding one solution v such that Tv = w, (not always possible); and
ii) computing the kernel of T.

In order to understand the kernel of a composition, we define a type of "inverse" that makes sense even for maps that are not injective.
Define: If T:V-->W is linear, and U is a subspace of W, the "inverse image" or "preimage" of U in V, is the set T^(-1)(U) = {all those x in V such that Tx is in U}.

Warning: This is confusing because we write also T^(-1) for the inverse of an isomorphism T. But every T has inverse images, even if it does not have an inverse map. Fortunately, if T is an isomorphism, and does have an inverse map called T^(-1), then T^(-1)(U) will be the same subspace in both meanings. The point here is that even if there is no map called T^(-1), there is still a subspace called T^(-1)(U).

We also define the inverse image of a point or a set, rather than a subspace, e.g. T^(-1)(w) = {all x such that Tx = w}, but such a preimage can be empty if w ≠ O.

Note: Inverse images generalize kernels, since ker(T) = T^(-1)({O}) is the inverse image of the zero subspace. We also write this as T^(-1)(O).

Here is the essential point:
Lemma: If T:V-->V is linear, and T^2:V-->V is (ToT) i.e. T composed with itself, then ker(T^2) = T^(-1)(ker(T)).
proof: To say T^2(x) = T(Tx) = O, means exactly that Tx is in ker(T), i.e. x is in T^(-1)(ker(T)). **QED.**

Remark: Similarly, if n > 1, then ker(T^n) = T^(-1)(ker(T^(n-1)).

We compute the kernel of P(D) in the case where T = P(D):S(R)-->S(R) is a polynomial in D that factors completely into linear factors: P(D) = (D-c1)(...)(D-cn). Since P(D) is a composition of the simpler operators (D-ci), the next proposition reduces the computation of the dimension of its kernel to that of the individual factors. It is absolutely basic to all work with linear maps, and must be thoroughly mastered.

Proposition: (fundamental theorem of linear algebra: **FTLA**) If T:V-->W is linear with finite dimensional kernel, and Y is a finite dimensional subspace of W, then T^(-1)(Y) is finite dimensional in V, and its dimension is at most dimY + dimkerT. Moreover equality holds if and only if Y is in the image of T.

proof: Choose a basis z1,...,zm of (Y meet ImT), and a basis x1,..,xn of kerT, and choose elements y1,...,ym such that T(yi) = zi for i=1,..,m. Claim: (x1,...,xn,y1,...,ym) is a basis for T^(-1)(Y). To see they span, let Tu belong to Y. Then it is in the image of T, so Tu = a1z1+...+amzm , for some ai. Then T(a1x1+...+amxm) = Tu also, so (a1x1+...+amxm – u) is in ker T hence (a1x1+...+amxm – u) = b1y1+...+bnyn for some bj, and hence u = (a1x1+...+am xm) –(b1y1+...+bnyn).

To see independence, assume (a1x1+...+am xm) –(b1y1+...+bnyn) = 0, for some ai,bj, so (a1x1+...+am xm) = (b1y1+...+bnyn). Apply T to get 0 = (b1z1+...+bn zn), whence all the bj = 0. But then (a1x1+...+am xm) = 0, and all the ai = 0. Since we have proved dimT^(-1)(Y) = dimkerT + dim(ImT meet Y), hence dimT^(-1)(Y) ≤ dimkerT + dim(Y), and equality holds if and only if Y is in the image of T, as claimed. **QED.**

Cor: Ker(D-a)^n+1:S(R)-->S(R) = (e^at) Pn.
proof: Since ker(D-a) has dimension one, it follows from the fundamental theorem above that dim Ker(D-a)^n+1 is at most n+1 dimensional. Since Pn has dimension n+1, so does (e^at) Pn. Since one checks immediately that this space is contained in the kernel of (D-a)^n+1, it equals that kernel. **QED.**

E.g.: The matrix of D on ker(D-a)^5 = (e^at) P4, with basis {(e^at)(t^r/r!)},0≤r≤4, is:

$$\begin{vmatrix} a & 1 & 0 & 0 & 0 \\ 0 & a & 1 & 0 & 0 \\ 0 & 0 & a & 1 & 0 \\ 0 & 0 & 0 & a & 1 \\ 0 & 0 & 0 & 0 & a \end{vmatrix}$$

, a classic example of a stretch plus a shift.

Remark: Im(D-a) = S(R), in fact we know ImD = S(R), so if f is in S(R) and Dg = fe^(-at), then (D-a)(g e^at) = f.
Cor: (D-a)^(-1)((e^at) Pn-1) = (e^at) D^(-1)(Pn-1) = (e^at) Pn.

Cor: Since ker(D-a) has dimension one, for the composition P(D), kerP(D) has dimension at most equal to degree P.

Claim: In fact kerP(D) has dimension precisely equal to degree P
proof: Case one: P(D) = (D-c1)....(D-cn) all ci different. Then {e^c1t,...,e^cnt} belong to kerP(D) and are independent. Thus they are a basis of kerP(D) in which the matrix of D is diagonal, with c1,...,cn on the diagonal.
Case two: P(D) = (D-c)^n+1. Then we have seen Ker(D-a)^n+1 = (e^at) Pn, and the matrix of D is a stretch plus a shift.
General case: P(D) = ∏(D-ci)^ni+1. Then kerP(D) = ∏ker(D-ci)^ni+1. Here the matrix of D consists of blocks which are [diagonal or] stretch plus shift. **QED.**
We will see that these matrices are typical for an operator T acting on ker(P(T)).

Inverting differential operators "locally"
Since we know the kernels of these operators, we see ker(D-1) meets P3 only in zero, so the map (D-1):P3-->P3 is invertible. In particular we can solve a non homogeneous differential equation like (D-1)y = 1+t^2, or y' – y = 1+t^2. E.g. we could introduce

bases and reduce the resulting matrix, but it is fun to see how to find an inverse operator more easily. Recall the famous geometric series 1/(1-x) = 1+x+x^2+x^3+...... In a sense this says that the inverse of (1-x) is the geometric series on the right. Now in case of an operator like D such that D^4 = 0 on the space P3, this series is finite, so we have that on P3, (1-D)(1+D+D^2+D^3) = 1. Hence (1+D+D^2+D^3) is the inverse (of 1-D), so the negative of this series inverts (D-1). So to solve (D-1)y = 1+t^2, apply (-1-D-D^2-D^3) to 1+t^2. Thus the solution is y = -(1+t^2) – 2t -2 = -t^2 -2t -3, as one can check. Adding to this the general solution of (D-1)y = 0, which we know is ce^t, gives the general smooth solution of (D-1)y = 1+t^2, namely y = -3 -2t -t^2 + ce^t. By composing inverses of operators like (D-c) = -c(1 – D/c), this trick inverts any operator P(D) on an appropriate space when the polynomial P(t) splits into linear factors, e.g. over C.

Minimal polynomials and cyclic subspaces

We have seen how to find natural bases for the solution spaces P(D) = 0, i.e. for V = ker(P(D)), where D is the differentiation operator on smooth functions and P is a polynomial that factors completely into linear factors. Moreover we have seen that the matrix of D in these natural bases is composed of blocks having a very simple form: namely to each factor of form (t-c)^k of our polynomial P(t), there is a k by k block having c's along the diagonal, and 1's just above the diagonal. In fact, this has little to do with the nature of the operator D or the fact that it is defined on a space composed of smooth functions. Rather it has to do with the fact that the space V our operator D is acting on, is the kernel of a polynomial in D, which has all its roots in k.

In fact for every operator T on a finite dimensional space, every vector is in the kernel of some polynomial in T. I.e. let x be any non vector in a space V, and T a linear operator on V. If V is finite dimensional, then in the sequence (x,Tx,T^2x,...,T^nx,.....) there is a smallest n such that the sequence (x,Tx,T^2x,...,T^nx) is dependent. Thus T^n + an-1T^n-1x +....+a1Tx + a0x = 0, for some unique sequence of scalars (a0,...,an-1) in k. The monic polynomial t^n + an-1t^n-1x +....+a1tx + a0, is called the **minimal polynomial** of x with respect to T.

The following corollary of the division theorem for polynomials is fundamental.
Lemma: If f(t) is the minimal polynomial of x ≠ 0 in V, w.r.t. T:V-->V, and g(t) is any polynomial in k[t] such that g(T)(x) = 0, then f divides g in k[t].
proof: By the division theorem for polynomials, there are polynomials q,r in k[t] such that g = qf + r, and where either r = 0, or deg(r) < deg(f). Substituting T into this equation, 0 = g(T) = q(T)f(T) + r(T). Since f(T) = 0, thus also r(T) = 0. By definition of f, no non zero polynomial of lower degree in T annihilates x, so r = 0. **QED.**

Corollary: If T:V-->V is linear, and (x1,...,xn) is a basis of V, the least common multiple m(t) of the minimal polynomials of the xi, is the unique monic polynomial of lowest degree in T that annihilates V. Any polynomial that annihilates V is a multiple of m(t).
proof: If fi is the minimal polynomial of xi, and g = h.fi, then g(T)(xi) = h(T)fi(T)(xi) = h(T)(0) = 0. Thus any multiple of fi annihilates xi, so any common multiple of the fi annihilates all xi, and thus also V. Any polynomial g which annihilates V, must annihilate all xi, and by the previous lemma, hence g is a multiple of all fi. Thus the monic

polynomial m of lowest degree annihilating V is the least common multiple of the fi. Any polynomial annihilating V is a multiple of all the fi, hence a multiple of m. Since m cannot divide any other monic polynomial of the same degree, it is unique. **QED.**

Definition: If T:V-->V is a linear map on a finite dimensional space over a field k, the unique monic polynomial m(t) of lowest degree such that m(T) = 0 on V, is called the minimal polynomial of T (on V).

Lemma: The roots of the minimal polynomial m, are those c such that ker(T-c) ≠ {0}.
proof: If c is a root of m(t), then m(t) = (t-c)f(t), so if ker(T-c) = {0}, then m(T)x= 0 for all x in V implies f(T)x = 0 for all x in V, so m is not the minimal annihilator for T. Conversely, if c is not a root of m(t), then m(t) = (t-c)f(t) + r with r ≠ 0. Thus -r.Id = (T-c)f(T), so (-1/r)f(T) is the inverse of T-c, so ker(T-c) = {0}. **QED**

Remark: 1) It follows from the previous results that the minimal polynomial of T can be computed from any basis of V, by using row reduction to determine the minimal polynomials of the basis vectors, and then finding their lcm.
2) Another common way to introduce the minimal polynomial of T is to consider the map k[t]-->Hom(V,V) sending t to T, and f(t) to f(T). Since the space k[t] has infinite dimension and Hom(V,V) has finite dimension when V does, there is a non zero kernel to this map which consists of multiples of some unique monic polynomial m of lowest degree in that kernel. Since Hom(V,V) has dimension = n^2, where dimV = n, there is a non zero kernel already in the space of polynomials of degree n^2 or less, which can be computed by row reduction from the image matrices of a basis of that space of polynomials. In fact, there is a non zero element of degree n or less in the kernel.

Example: if the matrix of T is:

$$A = \begin{vmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 0 \end{vmatrix},$$

then starting from e1 = (1,0,0), we have A(e1) = (1,0,2), A^2(e1) = (5,2,2), A^3(e1) = (9,6,12). Putting these four vectors in as columns and row reducing yields the matrix

$$\begin{vmatrix} 1 & 1 & 5 & 9 \\ 0 & 2 & 2 & 12 \\ 0 & 0 & 2 & 6 \end{vmatrix}.$$

The first three columns are pivots hence independent, so the sequence (e1, Ae1, A^2e1) is independent, hence a basis for V = Q^3. A kernel element of this matrix ending in 1, is (9,-3,-3,1) which gives as minimal polynomial for e1: t^3 -3t^2 -3t +9. By commutativity of polynomials in A, this monic polynomial is the one of minimal degree which annihilates the given basis, hence it is the minimal polynomial for A on V. This situation deserves some terminology.

Definition: A subspace U of V is T - cyclic with generator x, if U is spanned by the sequence (x,Tx,....,T^n x,......). Then U = k[T]x consists of all P(T)x with P in k[T].

Lemma: A polynomial f(t) annihilates a T cyclic space U with generator x, if and only if f(t) annihilates x. In particular, the minimal polynomial of a T cyclic space equals that of a generator.
proof: Certainly to annihilate all of U a polynomial must annihilate x. Conversely, if f(T)(x) = 0, then by commutativity of polynomials in T, for all j, f(T)(T^jx) = T^j(f(T)(x)) = T^j(0) = 0. **QED.**

We will classify operators by the complexity of their minimal polynomials. The simpler the minimal polynomial, the simpler the matrix we will find for T.

Definition: Two n by n matrices A,B are called similar, or conjugate, if there is an invertible n by n matrix Q such that B = Q^-1AQ.

Remark: Our goal is to find one simple representative within each conjugacy class of n by n matrices. Our solution is called the (generalized) Jordan form of the matrix, possibly found first by Camille Jordan but these things are often not precisely documented. The classical Jordan form occurs in most books, but the generalized version we present as an alternative to “rational canonical” form does not. I noticed it while thinking about generalizing the proof from Shilov for classical Jordan form, and it occurs in the superb book on abstract algebra by Chi Han Sah.

Nilpotent operators
We begin in the simplest possible case, with minimal polynomial t. The only operator satisfying this polynomial is T = 0, and the only matrix representing it, in every basis, is the zero matrix, all entries zero.

Next simplest is an operator T with minimal polynomial t^d, where d>1. Such operators are called nilpotent, since some power is zero. The easiest case, is d = n = dimV. Then if (x1,...,xn) is a basis for V, t^n is the lcm of the minimal polynomials for the basis vectors xi, so some basis vector x must itself have minimal polynomial t^n. Then (x,Tx,T^2x,...,T^n-1 x) is a T - cyclic basis for V. In this basis the matrix [T] is

$$\begin{vmatrix} 0 & 0 & 0 \cdots\cdots & 0 \\ 1 & 0 & 0 \cdots\cdots & 0 \\ 0 & 1 & 0 & 0 \\ \cdots\cdots & & & 0 \\ 0 & 0 & \cdots\cdots 1 & 0 \end{vmatrix}$$

This is about as simple as we can hope for. This is the matrix of the derivative D on the space Pn-1 of polynomials of degree < n, with basis (t^(n-1)/(n-1)!, ...,t^2/2!, 1, 1).

Next we look at maps T whose minimal polynomials have form (t-c)^d. If d = 1, then T = c.Id, so in every basis the matrix is diagonal, with c's on the diagonal, as follows:

$$\begin{vmatrix} c & 0 & 0 & 0 & 0 \\ 0 & c & 0 & 0 & 0 \\ 0 & \ldots & & & 0 \\ 0 & 0 & 0 & 0 & c \end{vmatrix}$$

The next simplest case is d = n = dimV as before. Again, if the minimal polynomial is (t-c)^n, and (x1,...,xn) is a basis, then some basis vector x has minimal polynomial (t-c)^n. But now we have a choice of cyclic bases for V. Since the minimal polynomial of x has degree n, the vectors (x,Tx,....,T^n-1x) are independent, but so are the vectors (x,(T-c)x,...,(T-c)^n-1 x). I.e. we can choose between a T cyclic basis, and a (T-c) cyclic one. Let us compare the matrices to see which is simpler.

In the T cyclic basis, each vector after the first is the T image of the previous one, and the last vector, has image T^nx = -(T-c)^n x +T^nx. If we expand (t-c)^n by the binomial theorem, we get some polynomial t^n – n t^n-1c + n(n-1)/2 t^(n-2)c^2-.....+(-1)^n c^n = which we will just call t^n + an-1t^n-1 + ...+a1t + a0. So the matrix is as simple as before except for the last column. I.e. we get:

$$\begin{vmatrix} 0 & 0 \ldots\ldots & 0 & -a0 \\ 1 & 0 \ldots\ldots & 0 & -a1 \\ 0 & 1 \ldots\ldots & 0 & -a2 \\ 0 & 0 & & \\ \ldots & \ldots & \ldots & \\ 0 & 0 \ldots\ldots & 1 & -an-1 \end{vmatrix} = \text{Cf, called the "companion matrix" of f = (t-c)^n.}$$

If instead we use the (T-c) cyclic basis (x, (T-c)x,....,(T-c)^n-1 x), then the T image of each of the first n-1 basis vectors equals the next basis vector plus c times itself, and the T image of last basis vector T(T-c)^(n-1)x) = (T-c)(T-c)^(n-1)(x) + c(T-c)^n-1 x = 0 + c(T-c)^n-1 x, is just c times itself. So we get this simpler matrix:

$$\begin{vmatrix} c & 0 & 0 & 0 & 0 \\ 1 & c & 0 & 0 & 0 \\ 0 & 1 & c & 0 & 0 \\ 0 & \ldots & \ldots & c & 0 \\ 0 & 0 & 0 & 1 & c \end{vmatrix}. = \text{J(c),n, called the elementary n by n jordan block for c.}$$

Put another way, since T has minimal polynomial (t-c)^n, the operator T-c is nilpotent with minimal polynomial t^n. Thus in the (T-c) cyclic basis, the matrix of T-c looks like the previous nilpotent matrix, and T itself is just (T-c) + c.Id.

We call the matrix of T in the T cyclic basis, the **rational canonical form** of T, and the matrix of T in the (T-c) cyclic basis is called the **Jordan form** of T.

Definition: A map T is nilpotent of index n, if T^n = 0, but T^(n-1) ≠ 0. A non zero vector x has nilpotency index k>0 for T if T^k (x) = 0, but T^(k-1) (x) ≠ 0. Thus to say T is nilpotent of index n on V, means every vector x ≠ 0 in V has index ≤ n, and some vector x ≠ 0 has index exactly n.

Thus an n by n matrix of T is in elementary Jordan form for the given basis (x1,...,xn), if the basis is (T-c) - cyclic, where (T-c) is nilpotent of index n. Sometimes it is convenient to arrange the basis vectors in the opposite order. Then the Jordan matrix looks as follows, with the 1's above the diagonal.

$$\begin{vmatrix} c & 1 & 0 & 0 & 0 \\ 0 & c & 1 & 0 & 0 \\ 0 & 0 & c & 1 & 0 \\ 0 & 0 & 0 & c & 1 \\ 0 & 0 & 0 & 0 & c \end{vmatrix}.$$

Remark: I prefer to have the 1's below the diagonal, but they are above the diagonal in the case of the derivative operator with monomial basis ordered as usual, namely as (1, t, t^2/2!,....,t^(n-1)/(n-1)!). Perhaps this is why there is historical precedent for putting them above the diagonal.

The next theorem is our first big goal.

Theorem: If T:V-->V is linear, dimV finite, and the minimal polynomial m(t) splits into linear factors over k, m(t) = ∏(t-ci)^di, where all ci are distinct, then V is a direct sum of the subspaces Vi = ker(T-ci)^ni, and each Vi is itself a direct sum of (T-ci) - cyclic subspaces. Thus in a basis for V composed of unions of these (T-ci)-cyclic bases for the summands, the matrix of [T] consists of elementary Jordan blocks.

There are two steps in proving this theorem. The first is to decompose V into the sum of the Vi, and then the second is to decompose each Vi into a sum (T-ci) cyclic subspaces. The second step is harder, so we do the first one next. Notice even after decomposing, that in each (T-ci) cyclic subspace, one could choose instead to use a T cyclic basis, thus getting a "rational canonical" matrix made up of companion matrices.

Decomposing a space as a direct sum

If the minimal polynomial of T has more than one distinct irreducible factor, we will decompose V as a direct sum of subspaces on each of which T acts individually, and where each restricted minimal polynomial has only one irreducible factor, possibly repeated. Thus we reduce the analysis of T to the case where m(t) = (t-c)^d.

Definition: V is a direct sum of subspaces U,W, if U and W meet only in the zero vector, and V = U+W, i.e. every vector in V is a sum of a vector in U and a vector in W.

Exercise: V is a direct sum of U and W if and only if each vector in V is a sum of a unique pair of vectors, one in U, one in W.

A direct sum decomposition for V is the same as viewing V as a direct product of subspaces, and is achieved by dividing a basis into two disjoint subsets, as shown next.

Proposition: If V is finite dimensional, V is a direct sum of U and W if and only if

i) there is a basis for V of form (x1,...,xn; y1,...,ym) where (x1,...,xn) is a basis of U, and (y1,...,ym) is a basis for W;) if and only if
ii) for every basis (x1,...,xn) of U, and every basis (y1,...,ym) for W, the union (x1,...,xn; y1,...,ym) is a basis for V; if and only if
iii) the natural map UxW-->V taking (x,y) to x+y, is an isomorphism.
proof: (ii) If V is the direct sum of U and W, and (x1,...,xn) a basis of U, and (y1,...,ym) a basis for W, we claim the union (x1,...,xn; y1,...,ym) is a basis for V. Since every z in V is a sum z = x+y with x in U and y in W, certainly the set spans V. And if a1x1+...+anxn + b1y1+...+bmym = 0, then x = a1x1+...+anxn = -b1y1-....-bmym = y, where x is in U and y is in W. Thus x = y = 0, so all ai and all bj are zero, hence the set is also independent. This also implies (i). Moreover the natural map in iii) takes the basis (<x1,0>,...,<xn,0>;<0, y1>,...,<0,ym>) for UxW to the set (x1,...,xn; y1,...,ym), hence the map is an isomorphism if and only that set is a basis for V. Thus all three conditions are equivalent and all hold when V is a direct sum of U,W. Conversely, if say the map in iii) is an isomorphism, then the image of Ux{0}, i.e. U, meets the image of {0}xW, i.e. W, only in {0}, and U+W is the image of UxW, namely V. **QED.**

Definition: A subspace W of V, is said to be T invariant, where T:V-->V is a linear map, if T(W) is contained in W. Then T restricts on W to a map W-->W.

Cor: i) If U is a T invariant subspace of V, and (x1,...,xn; y1,...,ym) is a basis of V, with (x1,...,xn) a basis of U, then the matrix of T in this basis decomposes into 4 blocks, with zeroes in the lower left corner, and the matrix of T restricted to U in the upper left corner, as displayed here.

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

ii) If V is the direct sum of T-invariant subspaces U,W, with (x1,...,xn) a basis of U and (y1,...,ym) a basis for W , then in the basis (x1,...,xn; y1,...,ym) for V, the upper left corner n by n block is the matrix of T restricted to U-->U in the basis (x1,...,xn), and the lower right m by m block is the matrix for T restricted to W-->W in the basis (y1,...,ym). The upper right and lower left blocks are all zeroes, as below.

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

Remark: Similarly we can define a space V as a direct sum of more than two subspaces, say of U1,...,Ur, if every vector y in V is a sum y = x1+...+xr of a unique sequence of vectors x1,...,xr with each xi in Ui. As before, the uniqueness means that each Ui meets the sum of the other Uj only in {0}, and V is the direct sum of U1,...,Ur if and only if a basis for V can be obtained by taking the union of bases of the Ui; if and only if the natural sum map U1x...Ur-->V taking (x1,...,xr) to x1+...+xr, is an isomorphism.

If V is the direct sum of T invariant subspaces U1,...,Ur, the matrix of T in a basis of this type is composed of r blocks, along the diagonal, with each block being the matrix of T restricted to Ui-->Ui, in the basis for that subspace. Elsewhere there are zeroes.

To compute dimensions of spaces spanned by given vectors, we often use the following efficient principle (the "slick trick"):
Lemma: If for each i, there is a linear map Ti such that Ti(xj) = 0 for all j except j=i, then the sequence (x1,...,xn) is independent.
proof: If any xi does depend on the other {xj}j≠i, we would have Ti(xi) = 0. **QED.**

Exercise: If for each i, there is a linear map Ti such that Ti(xj) = 0 for all j < i, then the sequence (x1,...,xn) is independent; and the same holds if Ti(xj) = 0 for all j > i. More generally, if V is a sum of the subspaces Ui, and if for each i there is a map Ti:V-->W that annihilates all Uj except Ui, and is injective on Ui, then the sum is direct.

Lemma: If V is the direct sum of T invariant subspaces U and W, then the minimal polynomial of T on V is the lcm of the minimal polynomials of the two restrictions of T to U and to W.
proof: exercise**.**

Next we prove a general theorem that will reduce our analysis to the case of a minimal polynomial which has only one irreducible factor.
Relatively prime decomposition Theorem: Let T:V-->V be a linear map on a finite dimensional space over a field k, with minimal polynomial m = fg, where the f,g are relatively prime polynomials in k[t], and define U = ker(f(T)), W = ker(g(T)). Then U,W are T invariant, and V is the direct sum of U and W. The minimal polynomial of T on U is f, and the minimal polynomial of T on W is g.
proof: The subspaces U,W are T invariant, since if x is in kerf(T), then by commutativity of polynomials in T, also f(T)(Tx) = T(f(T)x) = T(0) = 0, so Tx is in ker(f(T)) as well.

Since f,g are relatively prime, we recall there exist polynomials A,B in k[t] such that Af + Bg = 1 = the gcd of f,g. Substituting T into this equation gives A(T)f(T) + B(T)g(T) = Id. Hence if x is in the kernel of both f(T) and g(T), then gives A(T)f(T)(x) + B(T)g(T)(x) = Id(x), where the left hand side = 0, hence also the right hand side = x=0. Thus the subspaces U,W meet only in zero.

On the other hand, if x is any vector in V, the same equation gives us x as a sum of the vectors A(T)f(T)(x) + B(T)g(T)(x). We claim the first of these is in kerg(T) = W, and the second is in kerf(T) = U. I.e. by commutativity of polynomials in T, g(T)A(T)f(T)(x) = A(T)(f(T)g(T)(x)) = A(T)(0) = 0, since (fg(T) annihilates everything in V. Similarly, B(T)g(T)(x) is in kerf(T) = U. Hence V = U+W, and the sum is direct.

The minimal polynomial of T on U must be some factor of f, and on W some factor of g. But the original minimal polynomial = fg, and is the lcm of the restricted ones. Hence the restricted minimal polynomials are f and g. **QED.**

Remark: By induction, if m(t) = ∏fi^ni, all fi distinct irreducible polynomials, V is a direct sum of the T invariant subspaces Vi = ker(fi(T)), and on each Vi, the minimal polynomial is fi.

Now we can enhance our two special cases of Jordan's theorem, linking the nature of the minimal polynomial of T with the structure of a basis for T.
Definition: If v is a non zero vector such that T(v) = cv, for some scalar c, then v is called an eigenvector for T. The scalar c is called the associated eigenvalue.

Remark: Since Tx = cx for some x ≠ 0 iff ker(T-c) ≠ {0}, the eigenvalues are precisely the roots in k of the minimal polynomial.

Remark: It is immediate from the definitions, that a basis (x1,...,xn) consists entirely of eigenvectors of T, with eigenvalues ci, if and only if the associated matrix for T is diagonal, having the entries ci along the diagonal, and 0's elsewhere. Notice this is a Jordan matrix in which all the elementary blocks are one by one, hence there are only the eigenvalues along the diagonal, and no 1's above or below the diagonal.

Diagonalization Theorem: If T:V-->V is a linear map on a finite dimensional space over a field k, with minimal polynomial m(t) = ∏(t-ci) with all ci distinct, then V has a basis consisting entirely of eigenvectors for T, hence in this basis the matrix for T is diagonal. Conversely, if T has a basis in which the matrix is diagonal, then the minimal polynomial m for T factors over k into distinct linear factors.
proof: By the decomposition theorem, V is the direct sum of the Ui = ker(T-ci). On Ui, this means T equals ci.Id, hence in any basis for Ui, the matrix for T is diagonal, with ci along the diagonal. Hence, combining these eigenbases gives a basis for V in which the matrix of T are diagonal blocks, and hence gives a diagonal matrix for T.

Conversely, if T has such a matrix, then on each subspace Ui spanned by the eigenvectors for the scalar ci, T satisfies the polynomial (t-ci). Hence the polynomial ∏(t-ci) annihilates the given basis for V, hence all of V. Thus the minimal polynomial m for T, must divide this one. But no product of a proper subset of these factors can annihilate all of V. E.g. if we omit the factor (t-ci) leaving only ∏j≠i (t-cj), then applying ∏j≠i (T-cj) to an eigenvector x for ci, gives ∏j≠i (ci-cj) x, a non zero multiple of x. **QED**

Remark: It is easy to find a diagonalizing basis in practice for such a matrix A, given the roots ci of m. I.e. for each i, we find a basis of ker(A-ci) by row reduction. Any basis for V which is a union of bases of these kernels is a diagonalizing basis for A.

The next case of Jordan's theorem generalizes the behavior of the derivative operator D.
Theorem: Assume T:V-->V is a linear map on an n dimensional space/k, and the minimal polynomial is m(t) = ∏ (t-ci)^ni, all ci distinct in k, with deg(m) = n = dimV. Then in some basis for V, the matrix for T is composed of elementary Jordan blocks, one block for each i, with ci along the diagonal of the ith block, which is ni by ni, with 1's just below (or above) the diagonal, and 0's elsewhere.
proof: This follows from our previous results. I.e. since the ci are distinct, the factors (t-ci)^ni are pairwise relatively prime, so V decomposes as a direct sum of the T invariant subpaces Ui = ker((T-ci)^ni), and on each Ui the minimal polynomial is (t-ci)^ni. Thus each subspace Ui has dimension at least ni, but since n is the sum of the ni, none can be

larger. Thus we are in the setting of the proof above for an operator T with minimal polynomial (t-ci)^ni where ni is the dimension of the space Ui.

Hence in each Ui, we can choose a (T-ci) - cyclic basis of form (xi, (T-ci)xi,....,(T-ci)^(ni-1)xi), [or we can take the opposite ordering ((T-ci)^(ni-1)xi,..., (T-ci)xi, xi)]. In such a basis, the restriction of T to Ui-->Ui has an elementary lower [or upper] Jordan matrix. Thus taking the union of such bases for the Ui gives the desired basis for V. **QED.**

Remark: A Jordan basis is also easy to find in this case from a matrix A, but more work. For each i, compute a basis of ker(A-ci)^ni-1, which will be ni-1 dimensional, and find any vector xi in ker(A-ci)^ni that is not in ker(A-ci)^ni-1. Then xi generates a (T-ci) cyclic basis (xi,(T-ci)(xi),....,(T-ci)^ni-1(xi)) for ker(A-ci)^ni. The union of these bases is the desired jordan basis for V.

Next we treat the general case of Jordan's theorem, with no conditions on the degree of the minimal polynomial or the multiplicity of the irreducible factors, just assuming it factors completely into linear factors over k. This version thus applies to all operators over an algebraically closed field like C. This proof is more work, and already the first two cases are quite useful.

Remark: After years of trying to understand this topic clearly in the general case, I finally read the lovely clear exposition in Shilov, where he renders it all transparent simply by arranging the basis vectors in a 2 dimensional diagram I had not seen before. I have tried to reproduce the clarity of this explanation below, and I refer you to his inexpensive and superb little Dover book (see top of p. 135) for his account. Insel, et al. use a similar diagram they call a "dot diagram".

I like to use the concept of quotient spaces, to study the way the kernels of the various powers (T-c)^r grow as r increases, so we review them next.

Quotient Spaces: every subspace is a kernel

We know how to "multiply" spaces together to make larger product spaces. We can also "divide" a space by a subspace, analogous to the construction of quotient rings by ideals such as Z/n. This will imply that every subspace is a kernel of some linear map.

The equivalence relation defined by a subspace

Exercise: Given a subspace W of a vector space V, define a relation on V by saying x and y are equivalent if and only if x-y belongs to W. Show this is an equivalence relation on V, and define V/W = the set of equivalence classes of this relation.

We claim this equivalence relation is respected by the linear operations on V. I.e. if x, x' are equivalent and y,y' are equivalent, then x+y should be equivalent to x'+y'. To see this, recall x+y and x'+y' are equivalent if (x+y) – (x'+y') belongs to W. But this equals (x – x') + (y - y'), which does belong to W, since both (x-x') and (y-y') do so. Similarly, cx and cx' are equivalent. Thus if [x], [y] denote the equivalence classes of x and of y, we can define [x] + [y] = [x+y]. I.e. it does not matter which representative of [x] we use for the addition, since if [x] = [x'], and [y] = [y'], we have just proved that also [x+y] = [x’+y’]. So addition is well defined. Similarly we can define c[x] = [cx], since if [x] = [x’], then as above [cx] = [cx’].

Another way to say this is that the natural projection map π:V-->V/W is linear. I.e. If π(x) = [x], then π(x+y) = [x+y] = [x] + [y] = π(x) + π(y). Since π is obviously surjective, it follows from our general correspondence principle, that with the operations just defined, the quotient space V/W also satisfies the axioms of a vector space.

Exercise: Every subspace W of a vector space V, is the kernel of some linear map, namely W = ker(π), where π:V-->V/W is the projection π(x) = [x].

Induced maps on Quotients

The basic fact that fibers of a linear map are translates of the kernel, says that fibers of T become single elements of V/ker(T).

Lemma: If T:V-->W is linear, then for any elements x,y in V, Tx = Ty if and only if [x] = [y] in V/ker(T). I.e. the equivalence class [x] in V/ker(T) consists of all elements of V having the same image under T as x. I.e. we could also say that x and y are equivalent in V/ker(T) if and only if T(x) = T(y).

Cor: If T:V-->W is linear, then T factors as T’π, for a unique linear map T’:V/ker(T)-->W, where π:V-->V/ker(T) is projection. Moreover T’ is always injective.

proof: Since [x] = [y] if and only if T(x) = T(y), we can define T’([x]) = T(x). Then T’([x]) = 0 if and only if T(x) = 0, if and only if [x] =[0] in V/ker(T). I.e. T’ is injective. **QED.**

Lemma: If V has finite dimension and W is a subspace, then dimV = dimW + dimV/W.

proof: Since π:V-->V/W is surjective with kernel W, this follows from the fundamental dimension result: FTLA. **QED.**

Since when T:V-->V has minimal polynomial of form ∏(t-ci)^ni, we can always decompose the space V into the direct sum of subspaces Vi where (T-ci) is nilpotent, the next lemma is the central result for Jordan’s general theorem.

Proposition (existence of a jordan basis for a nilpotent map): If T is a nilpotent linear map on a finite dimensional space V, with minimal polynomial t^d, there is a basis for V consisting of a union of T – cyclic sequences, i.e. of form (x1,T(x1),...,T^(d1-1)x1; x2, T(x2),...,T^(d2-1)(x2);....;xr, T(xr),...,T^(dr-1)(xr)). Moreover, d1 ≤ d2 ≤ ...≤ dr = d. I.e. there is at least one T cyclic sequence of length d, and the others have lengths ≤ d.

Cor(existence of jordan form for any map with split minimal polynomial):

(i) If T:V-->V is linear on a finite dimensional V, with minimal polynomial m(t) = ∏(t-ci)^di, all ci distinct, then V is the direct sum of the subspaces Ui = ker(T-ci)^di, and each Ui has a basis which is a union of (T-ci) – cyclic sequences.

(ii) In the basis for V consisting of the union of these bases for the Ui, the matrix for T has a sequence of elementary Jordan blocks along its diagonal. For each i, the largest elementary block with ci on the diagonal has dimension di by di.

Proof of existence of nilpotent jordan form:

Assume T is a nilpotent linear map on a finite dimensional space V, with minimal polynomial t^d, define Kr = ker(T^r), and let (x1,...,xpd) be vectors in V representing a basis of Kd/Kd-1. I.e. they are a maximal set such that no non trivial linear combination of them belongs to Kd-1. Then note that T:Kd/Kd-1 --> Kd-1/Kd-2 is injective, since if T^d-2(Tx) = 0, then T^d-1(x) = 0. Thus (T(x1),...,T(xpd)) is independent in Kd-1/Kd-2, and we can extend the sequence to a basis (T(x1),...,T(xpd); y1,...,ypd-1) of Kd-1/Kd-2.

Continuing in this fashion, we reach the basis (T^(d-1)(x1),...,T^(d-1)(xpd); T^(d-2)(y1),..., T^(d-2)(yp d-1);.....;z1,...,zp1), for K1 = ker(T). Arranging these vectors in a rectangular diagram is very helpful:

x1,...............................,xpd;

T(x1),.......................,T(xpd); y1,................................,yp d-1;

....

(T^(d-1)(x1),...,T^(d-1)(xpd); T^(d-2)(y1),..., T^(d-2)(yp d-1);.................;z1,...,zp1),

It is clear that the vertical columns of this diagram are T cyclic sequences. It follows from the FTLA that they give a basis for V. I.e. since a basis of K1 plus a basis of K2/K1 gives a basis of K2, the bottom two rows give a basis of K2. Then since a basis of K2 plus a basis of K3/K2 gives a basis of K3, the bottom three rows give a basis of K3. Continuing, we see that the union of all the rows gives a basis of Kd = V. Thus we do have a basis for V that is a union of T cyclic sequences. **QED.**

Remark: This proves existence of the Jordan form, and uniqueness follows immediately, since pi is the number of elementary Jordan blocks of size i by i. Thus if ni = dimKi, then ni – ni-1 = mi = dim(Ki/Ki-1), and mi – (mi+1) = pi = number of i by i elementary blocks. Since the sequence of numbers p1,...,pd, determines the structure of the Jordan form, this is determined by the dimensions of the spaces Ki = ker(T^i), hence by T.

Remark: Finding Jordan bases in this general case is not so easy, but there is an algorithm for doing it, by following the steps of this existence proof. I.e. for each factor (t-c)^d in the minimal polynomial of the matrix A, look at the kernels of the spaces Kr = ker(A-c)^r, for all r ≤ d. What we want are bases for the quotient spaces Kr/Kr-1. One way to find them is to start from a basis for K1, then extend that to a basis for K2, then extend to a basis for K3,..., eventually getting a basis for Kd-1, which we extend to a basis for Kd by adding in the vectors x1,...,xpd. At each stage the newly added vectors form a basis for the quotient space.

In particular the vectors (x1,...,xpd) form a basis for Kd/Kd-1. Then as above we extend (A(x1),...,A^(xpd)) to a basis for Kd-1/Kd-2. We can do this because our previous computation of a basis for Kd-1 extended from a basis for Kd-2, gave us a basis for the quotient Kd-1/Kd-2. Placing the vectors (A(x1),...,A^(xpd)) in front of a basis for Kd-1/Kd-2, and eliminating vectors dependent on previous ones, starting from the left, extends the sequence (A(x1),...,A^(xpd)) to a new basis for the quotient Kd-1/Kd-2. Continuing this tedious job eventually gives a Jordan basis for V.

Example: We will find the Jordan form and a Jordan basis for the following matrix A =

$$\begin{vmatrix} 2 & -1 & 0 & 1 \\ 0 & 3 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & 3 \end{vmatrix}.$$

Using the theory of determinants, we calculate the characteristic polynomial, expanding along the first column, ch(t) = det(t-A) = (t-2)^3(t-3). Then the minimal polynomial m(t) divides this. This implies the minimal polynomial m(t) does split into linear factors and there is a Jordan form. To find it we only need the roots of m(t) which are the same as the roots of ch(t). So we just need to compute bases for the kernels of powers of (A-2) and (A-3). There is no mystery about ker(A-3), which must have one dimensional kernel since (t-3) occurs with exponent one in ch(t) hence also in m(t). It is easy to check (1,0,0,1) belongs to ker(A-3) hence is a basis.

Row reduction shows (A-2) has rank two, so ker(A-2) has dimension two, and one basis of K1 = ker(A-2), is (0,1,1,1), (1,0,0,0). We can extend this to a basis of K2 by adding the vector x = (0,0,-1,1), which additional vector thus gives a basis of K2/K1. Now remember we have to use this last vector ro find a new basis of K1. I.e. we compute (A-2)(0,0,-1,1) = (1,1,1,1), and extend to a new basis {(1,1,1,1), (1,0,0,0)} of K1. (All we have to do is add to (1,1,1,1) one of the old basis vectors for K1 that was not proportional to (1,1,1,1). So we chose the simpler one.) Then our basis of K2 is a union of two A cyclic sequences: {(0,0,-1,1), (1,1,1,1); (1,0,0,0)}, and our basis of ker(A-3) of course was (1,0,0,1). So our (lower) Jordan basis for A is the union of these four vectors: {(0,0,-1,1), (1,1,1,1); (1,0,0,0); (1,0,0,1)}. If Q is the matrix with these vectors as columns, i.e. Q =

$$\begin{vmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{vmatrix},$$

then we get the following Jordan form: Q^(-1)AQ = J =

$$\begin{vmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{vmatrix}.$$

End of example.

I find the following simple conceptual formulation of the theorem helpful.
Definition: Given T:V-->V, a T invariant subspace W of V is <u>indecomposable</u>, if it is not the direct sum of two non zero T invariant subspaces.

The following result is equivalent to the previous Proposition for nilpotent Jordan form.
Proposition: If T:V-->V is linear and nilpotent on a finite dimensional space, the indecomposable T invariant subspaces, are precisely the T - cyclic subspaces.
proof: Since we know every space decomposes into T cyclic subspaces, it suffices to show a T cyclic space is indecomposable. But a T cyclic space of dimension has minimal polynomial t^n. If it were decomposable, the minimal polynomial t^m, with m < n, of the largest summand would annihilate the whole space, a contradiction. **QED.**

Remark: If we prove this proposition independently, the previous Proposition will follow, since by the well ordering principle, a finite dimensional V must decompose into indecomposable subspaces. If those spaces are T cyclic, the Jordan structure theorem would follow.

Digression: We will give an idea of how to prove directly all indecomposable subspaces are T cyclic, by showing how to decompose a non cyclic one. Assume T is nilpotent.

Step One: Assume V is spanned by just two cyclic sequences i.e. V = Span(x, Tx,...,T^n-1x; y,Ty,...,T^m-1y). Although each cyclic sequence is an independent sequence, we allow that their union may not be independent. Assuming the minimal polynomials of x,y are t^n and t^m respectively, where n ≥ m, we will show that one can choose a new y' = y – P(T)x, for some polynomial P, such that the union of the T cyclic sequences generated by x and y' is independent, hence a basis for V.

To see this, let X = k[T]x = Span(x,Tx,...,T^n-1x), and consider V/X, which has dimension s ≤ m ≤ n. (Of course if dimV/X = m then by FTLA (x, Tx,...,T^n-1x; y,Ty,...,T^m-1y) is already a basis for V.) Now there is some smallest r such that T^r(y) belongs to X, and then by the division algorithm R(T)(y) belongs to X for a polynomial R iff t^r divides R(t). Hence (x,Tx,...,T^r-1x) is independent mod X, and all higher powers T^r+a (y) belong to X. Hence (x,Tx,...,T^r-1x) is basis of V/X so r = s.

Now since T^s(y) is in X = k[T]x, there is some polynomial Q(t) such that Q(T)x = T^s(y). Then 0 = T^m(y) = T^(m-s)T^s(y) = T^(m-s)Q(T)x. Thus t^(m-s)Q(t) is divisible by the minimal polynomial t^n of x, and hence Q(t) is divisible by t^(n-m+s). By hypothesis n ≥ m, so t^s divides Q(t), and Q(T)x = T^s P(T)x, for some polynomial P. Thus T^s(y) = Q(T)x = T^s P(T)x, so 0 = T^s(y)-T^s P(T)x = T^s(y-P(T)x). If we take y' = y-P(T)x, then T^s(y') = 0, and for all a, T^a(y) = T^a(y') mod X. Thus the sequence (y',Ty',...,T^(s-1)y') is a basis for V/X, and (x,Tx,...,T^n(x); y',Ty',.....,T^s-1(y')) is a basis for V which is a union of two cyclic sequences, as desired.

Step two: Now we want to do induction on dimension. We have to be a little careful in the order of arguments. Let x be an element with minimal polynomial t^n with n maximal, and put X = k[T]x. Then V/X has degree less than V, and since X is T invariant, the map T descends to T:V/X-->V/X. By induction there is a basis of V/X consisting of T cyclic sequences (y,Ty,...,T^m-1(y); ....; z,Tz,.....,T^s-1(z)). Of course this means only that T^m(y) = 0 =.....= T^s(z) in V/X, i.e. that T^m(y),...,T^s(z) belong to X. But now we apply step one to each of the sets (x,Tx,...,T^n-1(x); y,Ty,...,T^m-1(y)), ..., (x,Tx,...,T^n(x); z,Tz,...,T^s(z)), replacing y,...,z by y',..,z' so that T^m(y') =....= T^s(z') = 0 in V. Then the union of T cyclic sequences (x,Tx,...,T^n(x); y',Ty',....,T^m(y');.......; z',Tz',....,T^s-1(z')) is a basis of V. **QED.**

Remark: This proof seems less helpful for finding an explicit Jordan basis.

Notice that for general T with split minimal polynomial, the decomposition of V is always into indecomposable T invariant summands, but we have a choice of bases for those indecomposable summands. If we choose the (T-ci) cyclic bases we get the Jordan form composed of Jordan blocks, and if we choose the T cyclic bases we get the rational canonical form composed of companion matrix blocks.

Sometimes the following statement is useful.
Prop: If the minimal polynomial of T:V-->V splits into linear factors/k, there are unique maps D:V-->V and N:V-->V such that T = D+N, D is diagonalizable, and N is nilpotent. Both D, N belong to the ring k[T] of polynomials in T, hence DN = ND.
Remark: The proof of the relatively prime decomposition theorem expressed the identity as a sum of projections pi on the subspaces ker(t-ci)^di, and each pi as a polynomial in T. Then D = Sum ci pi is a linear combination of these projections, and a polynomial in T.

Generalized Jordan form (rational canonical form)

We have understood the decomposition of a space with respect to an operator with split minimal polynomial. This is fine over an algebraically closed field like C, but there are many fields such as Q, R, and Z/p, where polynomials do not always split. For these we simply generalize the previous discussion, replacing linear factors (t-c) by arbitrary irreducible ones f(t). Fortunately the whole theory goes through in exactly the same steps, with almost the same proofs.

We can state the more general result as follows:
General decomposition theorem: If T:V-->V is any linear operator on a finite dimensional space V over any field k, then V is a direct sum of indecomposable T invariant subspaces. An invariant subspace U is indecomposable if and only if:
i) U is T cyclic and **ii)** the minimal polynomial of T on U is a power of an irreducible polynomial f over k.

Then we obtain standard matrices for T from those on a T cyclic subspace U. Again there are two common choices, the companion matrix of the minimal polynomial on U associated to a T cyclic basis for U, and a generalized Jordan matrix associated to a basis obtained from an f(T) cyclic sequence, where f^d is the minimal polynomial on U. The first is called the rational canonical form in many books, and the second could be called the generalized Jordan form. As in the cases above with minimal polynomial of form (t-c)^d, it seems to me that the Jordan form is simpler, hence preferable.

So we will prove a generalization of the Jordan form, assuming now that the minimal polynomial has form ∏fi^di where each fi is a different irreducible polynomial over k. Everything goes through as before, in appropriate formulation. Again, we give the proof of the easier cases first, since it never hurts to make it as easy on ourselves as possible. You may skip ahead to the general case, but the next few preliminaries are common to all cases, and should be read by everyone.

Companion matrix associated to a polynomial

Since the irreducible polynomial f(t) of degree n takes the place of the irreducible factor (t-c), in a generalized Jordan matrix we will have blocks of "irreducible" n by n cyclic matrices [Cf] corresponding to the one by one blocks [c] in a classical Jordan matrix. The matrix [Cf] corresponding to a polynomial f, irreducible or not, is called the companion matrix of f, and is defined as follows.

If f = a0 + a1t +...+an-1 t^(n-1) + t^n, is any polynomial of degree n, define the n by n companion matrix of f as the matrix [Cf] resembling the jordan elementary block for a nilpotent operator of minimal polynomial t^n, but instead of zeroes in the right hand column, it has the negative of the non leading coefficients of f, i.e. Cf is this matrix:

```
|0  0............0  -a0  |
|1  0 ...........0  -a1  |
|0  1 ...........0  -a2  |     =  Cf.
|0  0                    |
|......................... |
|0  0.......... 1  -an-1|
```

Thus if f = t^n, then Cf = C(t^n) is precisely the elementary Jordan block for a nilpotent operator of index n. We have mentioned these companion matrices above in case f = (t-c)^n, as an alternative to the elementary Jordan form. Now we will use them to generalize the diagonal entries of the elementary Jordan form. Recall, we know that if t^n is the minimal polynomial of T, then C(t^n) is the matrix of T on any n dimensional subspace with a T cyclic basis of form (x,Tx,...,T^n-1x). Exactly so, if f(t) = a0 + a1t +...+an-1 t^(n-1) + t^n, is the minimal polynomial of T, then on any n dimensional subspace U with a T cyclic basis of form (x,Tx,...,T^n-1x), the corresponding matrix of T on U is C(f). In fact the following lemma is easy to check.

Lemma: The following are equivalent for T:V-->V.
i) dimV = n, and the minimal polynomial of T is f = a0 + a1t +...+an-1 t^(n-1) + t^n.
ii) In some basis for V, T has matrix C(f).
iii) V has a basis: (x,Tx,...,T^n-1x), where T^n(x) = -a0x - a1T(x) -...-an-1 T^(n-1)(x).
proof: exercise.

Cor: A T invariant subspace U is T cyclic if and only if in some basis for U the matrix of T is a companion matrix, if and only if the minimal polynomial for T on U has degree equal to the dimension of U.

The matrix Cf is the analog of the scalar c along the diagonal of a classical Jordan block. So what is the analog of the 1's that are just below the diagonal of a classical Jordan block? In place of them, we will have the n by n matrix N, which consists entirely of zeroes except for a single '1' in the upper right corner. I.e. define N as the following matrix of the same dimensions as Cf:

```
|0  0  0  0  1|
|0  0  0  0  0|
|0  0  0  0  0|  = N.
|0  0  0  0  0|
|0  0  0  0  0|.
```

A generalized Jordan block for f, is composed of copies of Cf's and N's just as a classical Jordan block is composed of c's and 1's.

Definition: A generalized "d by d" elementary Jordan block for the irreducible polynomial f of degree n, is a (dn) by (dn) matrix, with d blocks equal to Cf along the diagonal, and below each block Cf, there is one copy of the matrix N.

Remark: Although we can define matrices Cf even when f is not irreducible, these matrices will arise in a generalized Jordan block only in case f is irreducible.
Thus we have the generalized (d.n) by (d.n) elementary Jordan block for f:

$$\begin{vmatrix} Cf & 0 & 0 & 0 & 0 & 0 \\ N & Cf & 0 & 0 & 0 & 0 \\ 0 & N & Cf & 0 & 0 & 0 \\ 0 & 0 & N & Cf & 0 & 0 \\ 0 & 0 & 0 & N & Cf & 0 \\ 0 & 0 & 0 & 0 & N & Cf \end{vmatrix} = J(f),d$$

In particular, when f(t) = (t-c), this is the old d by d elementary Jordan block for c:

$$\begin{vmatrix} c & 0 & 0 & 0 & 0 & 0 \\ 1 & c & 0 & 0 & 0 & 0 \\ 0 & 1 & c & 0 & 0 & 0 \\ 0 & 0 & 1 & c & 0 & 0 \\ 0 & 0 & 0 & 1 & c & 0 \\ 0 & 0 & 0 & 0 & 1 & c \end{vmatrix} = J(c),d$$

Now we are ready for the proofs. We can imitate the old proofs almost exactly, with one twist. Recall by the Euclidean algorithm that if f is irreducible, then k[t]/(f) is a field E. The twist consists in utilizing this field E.
Lemma: If k is a subfield of E, and V is a vector space over E, with E basis {xj}, and if {yi} is a k basis for E, then the set of pairwise products {yi xj} are a k basis for V.
proof: This lemma (from field extension theory) is an exercise in changing the order of a doubly indexed sum.

Next we have the analog of an eigenspace.
Proposition: Given T:V-->V, with dimV finite, if the minimal polynomial m = f(t) of T is irreducible of degree n, then dimV = rn is a multiple of n, and V has a basis in which the matrix of T consists of r copies of the companion matrix Cf along the diagonal.
proof: Since f is irreducible, f(T) annihilates V, so V is a vector space over the field E = k[t]/(f(t)), where we define the product of a vector x by a polynomial g(t) as g(T)(x). Since a k basis for V is an E spanning set, dimV is finite over E as well, and if we multiply each basis vector in an E basis for V, by the elements of a k basis for E, we get a k -basis for V. Thus if dimV = r over E, and (x1,...,xr) is an E basis for V, then since (1,t,t^2,...,t^n-1) is a k basis for E = k[t]/(f(t)), so a k basis for V is: (x1,T(x1),...,T^n-1(x1); ......; xr, T(xr),....,T^n-1(xr)). This basis is a union of T cyclic sequences of length n, and the corresponding matrix of T is the following, as claimed.

$$\begin{vmatrix} Cf & 0 & 0 & 0 & 0 & 0 \\ 0 & Cf & 0 & 0 & 0 & 0 \\ 0 & 0 & Cf & 0 & 0 & 0 \\ 0 & 0 & 0 & Cf & 0 & 0 \\ 0 & 0 & 0 & 0 & Cf & 0 \\ 0 & 0 & 0 & 0 & 0 & Cf \end{vmatrix}.$$ **QED.**

When combined with the relatively prime decomposition theorem, this yields an analog of the diagonalization theorem.

Corollary: Given T:V-->V with minimal polynomial m = ∏fi, where the fi are distinct irreducible factors, there is a basis for V in which the matrix for T consists of blocks along the diagonal, each block being the companion matrix C(fi) of one of the irreducible factors of m(t). Each factor fi occurs at least once. There are no matrices N.

proof: By the decomposition theorem, V is a direct sum of T invariant subspaces Vi, on each of which the minimal polynomial is fi, an irreducible polynomial. Each subspace has dimension > 0, or else the minimal polynomial would not involve that factor. Hence in each summand Vi the previous result gives a basis which is a union of T cyclic sequences of length ni = deg(fi), hence a matrix of the desired form. **QED.**

So the analog of a diagonal matrix in general is something like this:

```
|Cf1  0 |    0    0    0    0|
|0   Cf1|   0    0    0    0|
|0    0   |Cf2|   0    0    0|
......................
|0    0   ...    |Cfs   0    0|
|0    0   ...    |0   Cfs    0|
|...................................|
|0    0   ...    |0    0   Cfs|, where all the fi are irreducible/k.
```

Remark: i) This cannot be further refined, because each subspace spanned by one of the sequences (x,Tx,...,T^n-1x) in this basis, has no proper invariant subspaces. I.e. the analog of an eigenvector is a vector with irreducible annihilator. Such a vector spans a "simple" invariant subspace: one with no proper invariant subspaces, and the subspaces in this decomposition are all spanned by such vectors. Since the annihilator of every vector in a subspace divides the minimal polynomial for that subspace, if the minimal polynomial of a subspace is irreducible every vector in that subspace is a T generator.

ii) In the "diagonalizable" case, there are in each Vi as many general Jordan blocks as possible for fi and each block is as small as possible: deg(fi) by deg(fi). This is the case iff the minimal polynomial for T, is a product of distinct irreducible factors fi, each occurring with multiplicity one.

iii) To find a generalized Jordan basis in this "diagonalizable" case is not hard. If fi has degree ni, then for each i, start with any vector x1 in ker(fi(T)), and form the T cyclic sequence it generates (x1,Tx1,...,T^ni-1x1). Then choose any vector x2 in ker(fi(T)) which is not in the span of the sequence (x1,Tx1,...,T^ni-1x1), and add in the T cyclic sequence it generates, getting the sequence (x1,Tx1,...,T^ni-1x1; x2,Tx2,...,T^ni-1x2).

There is no problem with overlaps between the subspaces spanned by different vectors since the intersection of two distinct simple invariant subspaces is also invariant, hence zero. Continuing, we obtain a "diagonalizing" generalized Jordan basis, i.e. the matrix consists entirely of companion matrices of the irreducible polynomials fi.

The other easy case is in some sense the opposite, when there is only one general Jordan block for each irreducible factor. Again the minimal polynomial completely determines the matrix, and the corresponding basis is not hard to find.

Proposition: Given T:V-->V, with minimal polynomial m(t) = f^d, where f is irreducible of degree n, assume dimV/k = dn = deg(m(t)). Then in some basis, the matrix for T consists of exactly one generalized Jordan block for f, i.e. it has d copies of Cf along the diagonal, and d-1 copies of the matrix N defined earlier, one copy of N below each diagonal block Cf except the last.

Proof: Here we differ from some treatments. It follows from previous lemmas on minimal polynomials and cyclic sequences that there is some vector x in V for which f^d is the minimal polynomial. Then the T cyclic sequence generated by x is a T cyclic basis, in which the matrix for T is a companion matrix, but it is the companion matrix C(f^d) for f^d. This would be the rational canonical matrix for T. We want to choose the basis differently, as a generalized Jordan basis. Our basis will not be T cyclic, but will be constructed from an f(T) cyclic basis.

I.e. since f^d is minimal for T, it follows that f(T) is nilpotent on V with minimal polynomial t^d. Then there is an f(T) cyclic vector x in V generating a maximal f(T) cyclic sequence (x,f(T)x,....,f^d-1(T)x). In fact, thinking back to our analysis of nilpotent operators, if Kr = ker(f^r(T)), then with the current hypotheses, the quotient space Kd/Kd-1 is a non zero vector space over the field E = k[t]/(f(t), hence so is every other quotient Kr/Kr-1. Since each quotient has dimension ≥ one over E, it thus has dimension ≥ n over k. But since dimV = dn is the sum of the dimensions of these quotients, each space Kr/Kr-1 has k - dimension exactly n, and E – dimension exactly one.

Thus in the sequence (x,f(T)x,....,f^d-1(T)x), x is an E basis for Kd/Kd-1, Tx an E basis of Kd-1/Kd-2,...., and T^d-1x is an E basis for K1 = ker(f(T)). Hence by the lemma above relating k bases and E bases, a natural way to get a k-basis for each of these quotients, is to multiply each vector in an E basis, by the elements of a k basis for E. If deg(f) = n, then a k- basis for E = k[t]/f is just (1,t,t^2,...,t^n-1).

Thus we get this k basis for V:
(x,Tx,...,T^n-1x; f(T)x, Tf(T)x,..., T^n-1f(T)x; ....; f^d-1(T)x,...., T^n-1f^d-1(T) x).
In this basis for V, the matrix is indeed the generalized "d by d" elementary Jordan block for f; i.e. the (dn) by (dn) matrix with d copies of Cf along the diagonal, and d-1 copies of N just below it. **QED.**

Cor: Assume dimV is finite/k and the minimal polynomial m(t) of T:V-->V has degree(m(t)) = dimV. If the minimal polynomial m(t) = ∏fi^di, then in some basis for V, the matrix for T consists of one elementary Jordan block for each i. I.e. for each i, there is one "di by di" Jordan block J(fi),di on the diagonal.

proof: As before, this follows from the previous case by the general decomposition theorem. **QED.**

The lemma for the general Jordan theorem, when the minimal polynomial is a power of one irreducible factor, but can have degree less than dimV, is an exact analog of the lemma for nilpotent operators, with the same modifications we have already seen.

Theorem: If dimV is finite, and T:V-->V has minimal polynomial f^d, where f is irreducible over k, then in some basis, the matrix of T consists of blocks, each of which is an elementary generalized Jordan matrix, J(f),e, where e ≤ d, and the largest of these blocks is (at least one copy of) J(f),d.

proof: The proof is an exact analog of the old one, expanding on the proof just given for the special case. Again, instead of a union of T cyclic bases, we will show there is a generalized Jordan basis obtained from an f(T) - cyclic sequence, by the trick of viewing E = k[t]/f, as a vector space over k. I.e. our basis will be the smallest T cyclic set containing the given f(T) cyclic sequence.

We imitate the old proof almost exactly, with that one twist. I.e. since f is irreducible, again k[t]/f is a field E. Now however V itself is not a vector space over E, but the quotient spaces of interest in our proof are. I.e. since T has minimal polynomial f^d(t), it follows that f(T) is nilpotent with minimal polynomial t^d, so again define Kr = ker(f^r(T)). Note that f(T) maps Kr into Kr-1, i.e. if f^r(T)(x) = 0, then f^r-1(f(T)x) = 0, so Kr/Kr-1 is a vector space over the field E = k[t]/f, because f(T) annihilates Kr/Kr-1.

Then let (x1,...,xpd) be vectors in V representing a basis of Kd/Kd-1 as a vector space over E. I.e. they are a maximal set in V = Kd, such that no non trivial E- linear combination of them belongs to Kd-1. Then note that f(T):Kd/Kd-1 --> Kd-1/Kd-2 is injective, since if f^d-2(f(T)x) = 0, then f^d-1(T)(x) = 0. Thus (f(T)(x1),...,f(T)(xpd)) is independent/E in Kd-1/Kd-2, so extends to an E basis (f(T)(x1),...,f(T)(xpd); y1,...,ypd-1) of Kd-1/Kd-2. Continuing in this fashion, we reach the E basis (f^(d-1)(T)x1,...,f(d-1)(T)xpd; f^(d-2)(T)y1,..., f^(d-2)(T)yp d-1;.....;z1,...,zp1), for K1 = ker(f(T)). Arranging these vectors in a rectangular diagram gives:

x1,...............................,xpd;
f(T)x1,......................,f(T)xpd; y1,...............................,yp d-1;

....

(f^(d-1)(T)x1,...,f^(d-1)(T)xpd; f^(d-2)(T)y1,..., f^(d-2)(T)yp d-1;.................;z1,...,zp1),

It is clear that the vertical columns of this diagram are f(T) cyclic sequences. These columns do not give a basis for V over E, since V is not a vector space over E, because f(T) does not annihilate V. But we can pass to a basis over k for each of these quotient spaces. I.e. if we again multiply each basis vector in an E basis for a space, by the elements of a k basis for E, to get a k -basis for that space.

Assume f has degree n. If x is an element of V, multiplying it by the k basis (1,t,t^2,...,t^n-1) for E = k[t]/(f(t)), gives (x,Tx,T^2x,...,T^n-1 x). Thus enlarging the E basis (x1,.......,xpd) for Kd/Kd-1 in the diagram above to a k basis, gives the k basis (x1,Tx1,...T^n-1x1;.......;xpd, T(xpd),...,T^n-1(xpd)). Doing this all the way down, we get k - bases of each quotient space Kr/Kr-1, and hence by the FTLA as before, their union is a k basis for V.

Moreover, this basis breaks up into a union of bases for T cyclic subspaces as follows. Before enlarging the E basis, the first column was the following f(T) cyclic sequence of length d: (x1, f(T)x1, ..., f^d-1(T)x1). After enlargement, this became:

(x1,Tx1,...,T^n-1x1; f(T)x1,Tf(T)x1,..., T^n-1f(T)x1;... ;f^d-1(T)x1,...,T^n-1f^d-1(T)x1). This is not a T cyclic sequence, but is a basis for the cyclic space with T cyclic basis: (x1,Tx1,....,T^n-1(x1); T^n (x1), ...;T^2n(x1),... ;T^(d-1)n(x1),...,T^dn-1(x1)). Thus we have decomposed V as a direct sum of indecomposable T cyclic subspaces, namely the T cyclic subspaces generated by the columns of the diagram for f(T). **QED.**

Remark: If we choose instead, the actual T cyclic sequences at the end of the proof, as bases of the T cyclic subspaces decomposing V, we get a basis composed of T cyclic sequences, as in the classical (nilpotent) Jordan theorem. The resulting matrix would consist of blocks of companion matrices C(f^e) for powers of f.

Cor (generalized Jordan form): If T:V-->V is linear on a finite dimensional space/k, and minimal polynomial m = ∏fj^dj, with the fj distinct irreducible polynomials, then there is a basis for V such that the matrix for T consists of blocks, each block an elementary Jordan matrix J(fj),e for some power fj^e, with e ≤ dj. For each j, there is at least one maximal block J(fj),dj.

Remark: The generalized Jordan matrix for f^d with blocks of Cf's and N's contains more information than does the companion matrix C(f^d) for f^d, since it displays the irreducible factors f of f^d. I.e. J(f),d contains f, displayed d times, whereas C(f^d) only contains f^d multiplied out. This is analogous to factoring 729 as 3^6 and then discarding that information and displaying it as 729. Some books make the opposite choice and call the resulting matrix the rational canonical matrix for T. I cannot think of a reason for preferring that choice, in the present context, since our approach requires factoring the minimal polynomial into irreducible factors and the rational canonical form loses that factorization. The common point is that both correspond to the same direct sum decomposition of V into indecomposable T cyclic subspaces, but when the factor f occurs to a power higher than one, the generalized Jordan basis displays more information.

The method taught in more advanced courses of diagonalizing the "characteristic matrix" [t.Id –T] for T over the polynomial ring k[t], gives a procedure for finding a rational canonical form and simultaneously finding the minimal polynomial, and decomposes the space into cyclic but not indecomposable summands. The algorithm neither requires nor produces irreducible factors of the minimal polynomial, and uses only the Euclidean algorithm. This method yields a cyclic decomposition with the fewest possible number of cyclic summands. The minimal polynomial of each summand will have degree equal to the dimension of the summand, but the minimal polynomial will not necessarily be a power of an irreducible polynomial. Although more subtle to justify, the algorithm is simple in principle, and can be carried out without knowing the theory behind it.

Remark: In practice, it is not so easy to find a generalized Jordan basis since we must begin by choosing a basis for the quotient space Kd/Kd-1 as a vector space over the field E = k[t]/(f). Computations over that field are more complicated, but presumably can be done using the Euclidean algorithm in k[t]. Note that all the examples worked out in some books are in one of the two easy cases above, i.e. either the minimal polynomial is a product of distinct irreducible factors, or it has degree equal to the dimension of V.

Uniqueness: Uniqueness follows from combining our proof of existence with the uniqueness of the classical Jordan form. I.e. the generalized Jordan matrix for T is determined if we know the number of elementary “e by e” blocks for each factor fi^di in the minimal polynomial and each e ≤ di. But the proof showed the number of “e by e” blocks for fi equals the number of e by e blocks in the classical Jordan matrix for the nilpotent map fi(T). Since that number was determined by fi(T), which is itself determined by T, we have uniqueness.

The following conceptual statement is equivalent to the generalized Jordan theorem:
Theorem(general Jordan decomposition): Given T:V-->V, dimV finite, minimal polynomial m = ∏fj^dj, the indecomposable T-invariant subspaces of V are precisely those T cyclic subspaces on which the minimal polynomial is a power fj^e of some irreducible fj. V is a direct sum of such subspaces. For each j, the sequence of exponents occurring for the minimal polynomials of these subspaces is uniquely determined by T.

Cor: Each map T:V-->V has a matrix composed of a uniquely determined family of blocks, each of which is either the companion matrix or the elementary Jordan matrix, associated to a power of an irreducible polynomial.

Digression: As in the nilpotent case, we sketch an inductive proof that non cyclic subspaces are decomposable, thus implying the existence of a T cyclic decomposition. By the general decomposition theorem we may assume the minimal polynomial is a power of an irreducible polynomial f.
Step one: Let V contain x,y, let X = k[T]x and Y = k[T]y be the T invariant cyclic subspaces of V generated by x,y, and assume V = X+Y, not necessarily direct. As before we will replace y by an element y’ = y –P(T)x, so that the sum V = k[T]x + k[T]y’ is direct. Assume ann(x) = f^d is the monic polynomial of lowest degree in T that annihilates x, where f is irreducible in k[t], and ann(y) = f^e, where e ≤ d. As usual, since f^e(y) = 0 lies in X, the monic polynomial R(t) of lowest degree such that R(T)y is in X divides f^e hence equals f^s for s ≤ e. Since T induces a map on V/X making it T cyclic for y, with annihilator f^s, it follows that dim(V/X) = deg(f^s). We seek next a polynomial P(T) such that y-P(T)x has annihilator f^s.

We f^s(y) lies in X, f^s(y) = Q(T)x for some polynomial Q. Hence 0 = f^e(y) = f^(e-s)f^s(y) = f^(e-s)Q(T)x. Hence f^d, the minimal annihilator of x, divides f^(e-s)Q. Thus f^(d-e+s) divides Q, and d≥e imples f^s divides Q, so Q = f^sP for some P. Thus f^s(y) = Q(T)x = f^sP(T)x, so f^s(y-P(T)x) = 0, as desired. I.e. take y’ = y-P(T)x. Then V = k[T]x + k[T]y’ and dimV = dim(k[T]x) + dim(V/X) = dim(k[T]x) + dim(k[T]y’) implies then sum is direct.
Step two: This is exactly as before, assume the minimal polynomial of T on V is a power of an irreducible f, thatV = k[T]x + W, and by induction that V/(k[T]x) is a direct sum of T cyclic subspaces, k[T]y +...+k[T]z. Then apply step one to each subspace of V of form k[T]x+k[T]y,....,k[T]x+k[T]z, replacing y,....,z by y’,...,z’, and then the sum V = k[T]x + k[T]y’+...+k[T]z’ is direct by counting dimensions. **QED.**

Analogy with decomposing finite abelian groups as products of cyclic ones

Because this story is complicated, we want to tell a slightly simpler story that is completely analogous, of cyclic abelian groups and products of them. Remember Z/n from elementary number theory? This is a cyclic abelian group and a ring, where m = 0 iff n divides m. Other finite abelian groups are obtained from products of these. In fact such products give all finite abelian groups up to isomorphism, and the proof is analogous to that of existence of Jordan forms. This is because both proofs rest on the Euclidean algorithm, either in Z or in k[t]. Since abelian groups are easier to grasp, they can serve as useful tool for remembering what is true of linear maps.

So how are we to understand cyclic groups? E.g. Z/7 is a cyclic group, and the simplest property it has is its order, 7. Also multiplication by 7 kills all elements of the group, so 7 is both its order and its (minimal) annihilator. Moreover Z/7 has no non trivial subgroups because its order is prime. I.e. a non zero element of Z/7 is annihilated by a divisor of 7, hence only by 7. So for any non zero element x in Z/7, the cyclic sequence (x,2x,...,nx,...) it generates is all of Z/7.

Z/49 on the other hand is cyclic, but not every non zero element generates it, since some elements like 7, 14,... have order 7, while others like 1,8,.... have order 49. We can thus tell the difference between Z/49, and the product Z/7 x Z/7, which has order 49 but is annihilated by 7. So we can tell that Z/7 x Z/7 is not cyclic, because its annihilator is smaller than its order. Moreover, although Z/49 is not decomposable as a product, unlike Z/7 it does have a non trivial subgroup, generated by 7.

Now look at Z/7 x Z/8. This has order 56, but the element (1,1) is only annihilated by 56, since (n(1,1) = (n,n) is zero only when n = 0 mod 7 and also n = 0 mod 8, i.e. iff n = 0 mod 56. So in fact Z/7 x Z/8 ≈ Z/56 is really cyclic although it was presented to us as a product. This is called the Chinese remainder theorem (CRT).

Claim: Z/n x Z/m ≈ Z/(mn) iff n,m are relatively prime, iff the annihilator of Z/n x Z/m = nm, iff the annihilator equals the order.
Indeed a finite abelian group is cyclic iff its order equals its annihilator.

What does this have to do with linear operators? The annihilator of an abelian group corresponds to the minimal polynomial of an operator T. The order or cardinality of the group corresponds to the dimension of V. A better analogy is between the cardinality of G and the "characteristic polynomial" of T, whose degree equals the dimension of V. I.e. the annihilator of the group G divides the cardinality of G, just as the minimal polynomial divides the characteristic polynomial. For the theory of characteristic polynomials, see the later discussion on determinants. For now it is defined for a matrix A as the determinant of the polynomial matrix (tId – A), where tID is the diagonal matrix of the same dimensions as A, with t everywhere on the diagonal. The analog of the CRT is the relatively prime decomposition theorem for linear operators. A T invariant space U is T cyclic iff the minimal polynomial of T on U has degree equal to dimU. A T cyclic space has no proper invariant subspace iff its minimal polynomial is irreducible.

Just as every T invariant space is a direct sum of indecomposable T cyclic subspaces, every finite abelian group is a product of indecomposable cyclic groups, and knowing the cardinality and the annihilator of the original group can tell us something abut the number of factors of various orders. Say the group has order 210 = (2)(3)(5)(7). Then there must be cyclic factors annihilated by each of these factors, so there must be factors of Z/2, Z/3, Z/5, and Z/7. But the product of those factors already has order 210, so must exhaust the whole group. By the CRT the original group is also cyclic. I.e. if the order of an abelian group is a product of distinct prime factors then the group is cyclic and is a product of cyclic groups of those prime orders. The two extreme cases of Jordan decomposition have analogies here as well. If an abelian group has order p^n where p is prime, and is annihilated by p, then the group must be a product of n copies of Z/p. If the group G is annihilated by p^n which equals its order, then G = Z/p^n.

By analogy an operator T on a Q vector space V, of dimension 5 with minimal polynomial ch(t) = t(t-1)(t-2)(t^2+t+1) = a product of distinct irreducible factors over Q, must be T cyclic, and V must decompose into a direct sum of four indecomposable cyclic subspaces of dimensions 1,1,1,2.

If the annihilator of a finite abelian group is 5, then the group must be a product of copies of Z/5, and if the order of the group is say 125, then the group is Z/5 x Z/5 x Z/5. and that is the only possibility. This is analogous to the case of a linear operator with irreducible minimal polynomial f, i.e. the “diagonalizable” case. Then the matrix is a product of copies of C(f), and if dimV = dn where degree(f) = n, then the characteristic polynomial is f^d, and the Jordan form has d copies of C(f) along the diagonal.

So if a finite abelian group G has prime annihilator p, then G is a product of copies of Z/p, and the number of copies is found from the order of G. Analogously, if the minimal polynomial of T is an irreducible polynomial f, then the matrix of T is a product of copies of C(f), and the number of copies is computed from the dimension of V.

If the order of G is greater than its annihilator, there must be some prime p such that G has at least two factors such as Z/p^r x Z/p^s, in its decomposition. If the minimal polynomial m(t) of an operator has degree lower than dimV, there must also be some irreducible factor f of m(t) such that the general Jordan form has at least two cyclic blocks each with C(f) along the diagonal.

A finite abelian group of form: Z/p^d1 x ....x Z/p^dr with d1 ≤.....≤ dr, has cardinality ∏p^di and annihilator p^dr and each factor Z/p^di is not further decomposable, just as a direct sum of T cyclic subspaces Ui with minimal polynomials f^di where f is irreducible has characteristic polynomial ∏f^di, minimal polynomial f^dr, and each Ui is indecomposable.

We have emphasized the minimal polynomial and the dimension of our space, but have seen these do not always suffice to characterize the Jordan form. This is clear by analogy with the product decomposition of an abelian group. The group Z/3 x Z/3 x Z/9 has annihilator 9, and order 81, but so does the non isomorphic group Z/9 x Z/9. Only in extreme cases, like Z/3 x Z/3 x Z/3 x Z/3 with annihilator 3 and order 81, or Z/81 with order and annihilator 81, do these two pieces of data tell the whole story. This is analogous to determining the Jordan form from just the minimal polynomial and the dimension, where these extreme cases correspond to the first two “easy” cases of the Jordan form of T:V-->V, i.e. when the minimal polynomial m(t) was either irreducible or had degree equal to dimV.

To completely determine a finite abelian group G we need the whole sequence of orders of the factors of its product decomposition, just as we need the whole sequence of minimal polynomials of all the T cyclic subspaces in the decomposition of V. This sequence of prime powers for G, or of powers of irreducible polynomials for T, is called the sequence of elementary divisors. E.g. the elementary divisors (3,3,3^2,3^3) determine the group Z/3 x Z/3 x Z/9 x Z/27 up to isomorphism, just as the sequence (t,t, t^2, t^3) determines the operator T:k^7-->k^7 with matrix consisting of four blocks which are the companion matrices C(t), C(t), C(t^2), and C(t^3), up to conjugacy.

Cyclic versus indecomposable decompositions

This is a further source of complication. All indecomposable decompositions are cyclic, but not all cyclic decompositions are indecomposable. When the annihilator or the minimal polynomial has more than one irreducible factor, there is always a further decomposition, whether the object is already cyclic or not. On the other hand, if the minimal polynomial or annihilator has only one irreducible factor this is not the case. I.e. any cyclic decomposition of a group or map whose minimal annihilator is a power of one irreducible factor, is indecomposable. This means there are two natural choices of cyclic decompositions of a group and of a map, our “maximal” cyclic one into as many indecomposable cyclic pieces as possible, and a “minimal” one into as few cyclic pieces as possible. The maximal decomposition we have chosen carries its information more visibly, hence is presumably more useful.

To see the difference, Z/(210) is isomorphic by CRT to Z/2 x Z/3 x Z/5 x Z/7. If only a cyclic decomposition is needed, then Z/(210) is fine, but the indecomposable factors in the longer decomposition tell us more. The two are theoretically equivalent, since we can multiply (2)(3)(5)(7) together, or factor 210, but multiplying is much easier than factoring. In a non cyclic example, the group Z/3 x Z/3 x Z/2 x Z/4 is isomorphic to the group Z/6 x Z/12, since we can use the CRT to decompose this last group into the original four factors. This latest decomposition is in a sense minimal because since the group is not cyclic, it can not be decomposed with fewer than two cyclic factors. The last version is also unique in the sense that it is the only cyclic decomposition in which the orders of the successive factors divide each other.

This phenomenon occurs also for linear maps. E.g. the map T:k^3-->k^3 with minimal polynomial (t-1)t^2 has “maximal” i.e. indecomposable decomposition into two blocks, C(t^2) and C(t-1). But k^3 is already T cyclic since the minimal polynomial has degree 3 = dim(k^3), so there is also the possibility of using the one companion matrix C((t-1)t^2). The matrix for T composed of a minimal number of companion matrix blocks whose minimal polynomials divide each other, is what I call rational canonical form for T. The rational canonical form in some books is a compromise wherein the decompositions are maximal indecomposable cyclic subspaces, but the matrices on these subspaces are companion matrices C(f^d) of powers of irreducible polynomials f, rather than Jordan matrices containing d copies of C(f). It is like the difference between presenting an indecomposable factor group as Z/(729) as opposed to Z/(3^6).

Although there is less visible information in the “minimal” rational canonical form, it arises from a natural procedure. One can diagonalize the characteristic matrix (tId-A) of A using row and column operations over the Euclidean domain k[t]. This can be done so the polynomials on the diagonal divide each other, and these polynomials then give the minimal polynomials of the factors of this rational canonical decomposition of A. This is explained in the language of modules in notes on my website for math 8000 and math 845. The idea is a linear map T:V-->V allows us to view V as an abelian group with an action of the Euclidean ring k[t]. This gives a uniform way of presenting the discussion here. The connection is this: an operator T:V-->V focuses attention on the subring k[T] of L(V), and we have analyzed the action of k[T] on V. If m(t) is the minimal polynomial of T, then k[T] is isomorphic to k[t]/(m(t)), so it is more efficient to pull back and deal simply with the action of k[t] on V, via t.x = T(x).

Summary of facts:
Given a linear map T:V-->V, on V finite dimensional, define the following concepts:
i) A subspace U of V is T invariant, if T(U) is contained in U.
ii) A subspace U is T cyclic, and generated by x, if for some vector x, U is spanned by the sequence (x,Tx,T^2x,.....T^rx,......). Such subspaces are always T invariant.
iii) A T-invariant subspace U is simple if it has only the trivial T invariant subspaces {0} and U.
iv) A T-invariant subspace U is decomposable, if and only if there are non zero T invariant subspaces Ui and U2 such that U is the direct sum of U1and U2.

Then one has:(assume all groups are finite abelian, all subspaces T-invariant)
ia) The annihilator of a subgroup divides the annihilator of the group.
ib) The restriction of T to a T invariant subspace U has a minimal polynomial that divides the minimal polynomial of T.

iia) If G is the direct product of subgroups H,K, the cardinality of G is the product of those of H and K, and the annihilator of G is the lcm of those of H and K.
iib) If U is T invariant and a direct sum of invariant subspaces U1,U2, the characteristic polynomial on U is the product of the characteristic polynomials on U1 and U2, and the minimal polynomial on U is the lcm of the minimal polynomials on U1 and U2.

iiia) A group is cyclic iff its annihilator equals its cardinality.
iiib) A T- invariant subspace is T cyclic if and only if the minimal polynomial and characteristic polynomials are equal, if and only if the minimal polynomial has degree equal to the dimension of the subspace.

iva) If the ann(G) has more than one irreducible factor, G decomposes as a product.
ivb) If the minimal polynomial on a T invariant subspace U is not a power of an irreducible polynomial, then U is decomposable.

va) The annihilator of a group G divides the cardinality of G.
vb) The minimal polynomial of T has degree ≤ the dimension of the space and divides the characteristic polynomial.

via) If the annihilator is less than the cardinality of G, i.e. if G is not cyclic, then G decomposes as a product of subgroups.
vib) If the degree of the minimal polynomial of T is less than the dimension of the space, i.e. if the space is not T- cyclic, then it is decomposable.

viia) A group G is indecomposable iff it is cyclic of prime power order.
viib) A T -invariant subspace is indecomposable if and only if it is T-cyclic and its minimal polynomial is a power of an irreducible polynomial.

viiia) Every (finite abelian) group G decomposes as a direct product of indecomposable subgroups, i.e. of cyclic subgroups of prime power order. The sequence of prime power orders, called elementary divisors of G, is uniquely determined by G. The order of G is the product of all the elementary divisors, and the annihilator is the product of just the highest power of each distinct prime factor occurring as an elementary divisor.
viiib) Every finite dimensional space V with linear operator T, is a direct sum of indecomposable T invariant subspaces, i.e. T-cyclic subspaces on which the minimal polynomial is a power f^e of an irreducible polynomial, and such a subspace has dimension deg(f^e) = e.deg(f). The sequence of these powers of irreducible polynomials, called elementary divisors, is determined by T:V-->V. The characteristic polynomial is the product of the elementary divisors, and the minimal polynomial is the product of just the highest powers of each irreducible factor occurring as an elementary divisor.

We can imitate the proofs for Jordan decomposition, to decompose finite abelian groups.
Theorem: If a finite abelian group has annihilator ann(G) = ∏pi^ni, where the pi are distinct primes, and we define Gi = kernel of multiplication by pi^ni, then G is isomorphic to the direct product of the subgroups Gi.
proof: Use the Euclidean algorithm as before.

Exercise: If ann(G) = #(G) = n, then G is cyclic of order n, G = Z/n.
Exercise: If ann(G) = p is prime, then G = ∏ Z/p. [Hint: G is a Z/p vector space.]

Theorem: If ann(G) = p^d, with p prime, then G is a product of cyclic subgroups of orders p^e with e ≤ d, and at least one factor has order p^d.
proof: Exactly as before, define Kr = kernel of multiplication by p^r. Then each quotient group Kr/Kr-1 is a Z/p vector space and multiplication by p is a Z/p linear injection of Kr+1/Kr into Kr/Kr-1. Thus we can choose Z/p vector bases of each quotient group and arrange them in a diagram as before.
x1,...............................,xnd;
p(x1),......................,p(xnd); y1,...............................,yn d-1;

....

(p^(d-1)(x1),...,p^(d-1)(xnd); p^(d-2)(y1),..., p^(d-2)(yn d-1);.................;z1,...,zn1).

Here we are not concerned with choosing a basis, but merely with finding a decomposition of G into a product of cyclic subgroups. Recall that in the original proof of existence of classical Jordan form, each column represented a basis for a single T cyclic factor. Here each column of our diagram generates a single cyclic factor of our product of groups. Thus for each r, there are n(sub)r factors isomorphic to Z/p^r, and at least one factor isomorphic to Z/p^d. The reader will benefit from showing these cyclic factors give a direct product decomposition of G. **QED.**

A non trivial example: G = GL(3,Z/2) = invertible 3 by 3 matrices over Z/2
The so called collineation group of the 7 point projective plane is isomorphic to the group of invertible 3 by 3 matrices over the field Z/2. This makes it possible to determine all elements of this group explicitly and calculate their conjugacy classes by determining their generalized Jordan forms. This is a big advantage over using pure geometry as the reader may test for himself by trying to visualize say a collineation of order 7.

First we calculate the order of the group G. The first column can be any non zero vector in (Z/2)^3 and there are 7 of these. The second column must be a vector not on the line through the first one so there are 6 of those. Finally the third column is a vector not in the plane spanned by the first two and there are 4 of those, so #G = 168.

Next we compute one representative in each conjugacy class. These correspond to the possible general Jordan forms, hence to the possible sequences of elementary divisors. This must be a sequence of powers of irreducible polynomials such that the product of the polynomials in the sequence has degree 3. They also must not be divisible by t, since the minimal polynomial has 0 as root iff the matrix is not invertible. The only candidate irreducible polynomials over Z/2 are: (1+t), (1+t+t^2), (1+t+t^3), (1+t^2+t^3).

Thus the possible sequences of elementary divisors are:
A: (1+t, 1+t, 1+t); B: (1+t, (1+t)^2); C: (1+t)^3;
D: (1+t, 1+t+t^2); E: (1+t+t^3); F: (1+t^2+t^3).
I.e. there are exactly 6 conjugacy classes of elements in G.

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = A$$ has order one of course, and no other conjugates.

We use the rational canonical form of B since it has more zeroes than the Jordan.

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = B$$ has order two. To compute its conjugates we ask what it commutes with.

Since B is its own inverse, B commutes with another matrix X iff BXB = X. But,

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix}\begin{vmatrix} a & b & c \\ x & y & z \\ u & v & w \end{vmatrix}\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} a & b & c \\ x & y & z \\ u & v & w \end{vmatrix}$$ iff y = a, x = b, z = c, v = u, w arbitrary.

Thus if we let (a,b,c,u,w) be arbitrary, then (x,y,z,v) are determined. So there are 2^5 = 32 such matrices X. Setting detX = 1, gives w = 1 = a+b. So now once (a,c,u) are chosen, X is determined. There are 2^3 = 8 such X, so B has 168/8 = 21 conjugates, all of order 2. Similar calculations yield the following results.

The Jordan form of C is:

$$\begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = C,$$ which has order 4, and 42 conjugates, hence also of order 4.

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = D, \text{ of order 3, with 56 conjugates.}$$

$$\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} = E, \text{ order 7, with 24 conjugates.}$$

$$\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{vmatrix} = F, \text{ order 7 with 24 conjugates.}$$

Adding these numbers gives 1 + 21 + 42 + 56 + 48 = 168, the full group G. We can also check that the group generated by one element of order 7 contains elements of the other conjugacy class of elements of order 7. Recall a subgroup is called normal if it is invariant under conjugation by every element. Hence a normal subgroup of G which contains an element of any given order, contains all of them, thus has order = 1 + a sum of some of the numbers 21, 42, 56, 48. Since the order of a subgroup of G must divide 168, the only normal subgroups thus have order 1 or 168, i.e. G is a "simple" group. This group is interesting historically since it was missed by Jordan in his original attempt to classify all small simple groups in his book, Traite' des substitutions. It was discovered and described in detail by Felix Klein. It also arises as the automorphism group of the complex projective plane curve X^3 Y + Y^3 Z + Z^3 X = 0.

A numerical example

$$\begin{vmatrix} 0 & -7 & 14 & -6 \\ 1 & -4 & 6 & -3 \\ 0 & 0 & -4 & 9 \\ 0 & -4 & 11 & -5 \end{vmatrix} = A$$

We use the characteristic polynomial to determine the Jordan form. In this case, ch(t) = det(t-A) = (1+t^2)^2. Hence over Q, the minimal polynomial m(t) = either (1+t^2) or (1+t^2)^2. Computing A^2 + Id, shows it is not zero, so m(t) = ch(t) = (1+t^2)^2. this means the space Q^4 is already A cyclic and indecomposable. The Jordan form is this:

$$\begin{vmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{vmatrix} = J(1+t^2; 2). \text{ Note the matrix } N = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} \text{ in the lower left corner.}$$

The rational canonical form is this:

$$\begin{vmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{vmatrix} = C((1+t^2)^2).$$

Spectral theory

The previous discussion applies to all fields, e.g. the rational field, any finite field such as Z/p, or any field of form k[X]/(f) where k is a field and f is irreducible in k[X]. Now we specialize to two of our most familiar fields from analysis, the real and complex numbers. Is there some way we can deduce more about the structure of our matrix from our knowledge of those fields? I.e. we know C is algebraically closed, so every polynomial over C does split into (not necessarily distinct) linear factors. R is ordered, so it makes sense to say whether a number is positive. Moreover the fundamental theorem of algebra also implies that the only irreducible polynomials over R are either linear or quadratic. I.e. if f has real coefficients, and z is a complex root, then zbar is also a complex root, so (x-z)(x-zbar) is a real irreducible quadratic factor of f. Thus the non real complex linear factors pair up two at a time to give irreducible real quadratic factors.

If A is a complex n by n matrix, its minimal polynomial is a product of linear factors, possibly with multiplicities, m = ∏(t-ci)^ni, where the ci are complex numbers. If A is a real n by n matrix, its minimal polynomial is m = ∏(t-ci)^ni ∏qj^mj, where the ci are real numbers, and the qj are irreducible monic real quadratic polynomials. The irreducible real quadratic factor q(t) = (t^2 - 2at +(a^2+b^2)) = (t-z)(t-zbar), where z = a+b sqrt(-1). We will prove some theorems that tell us more, such as when there are no multiplicities, i.e. when all the exponents ni = mj = 1, or when there are no quadratic factors i.e. when all mj = 0.

The fundamental concern is always to decompose our space into smaller subspaces on which the operator acts more simply, and the idea here is to use orthogonal decompositions. For that the basic tool is the dot product, which measures angles, lengths and perpendicularity, at least in R^n. If we assume some link between A and the dot product, e.g. that A is length preserving, we may be able to deduce a decomposition theorem by showing that orthocomplements of A invariant subspaces are also A invariant. All our spaces will have finite dimension.

Real dot products

If x,y are vectors in R^n, we write [x],[y] for the column vectors they represent. Then the dot product is defined by setting x.y = [x]^t [y]. I.e. if [x] is a column vector, then [x]^t is a row vector, so we can multiply [x]^t by [y] as matrices. Writing out the coordinates x = (x1,..,xn), and y = (y1,...,yn), this is the usual definition from calculus: x.y = x1y1+...+xnyn. Then the following properties hold, for x, x',y in R^n, c in R:

1) x.y = y.x, (symmetry).
2) (cx).y = c(x.y), (x.cy) = c(x.y), (bilinearity).
3) (x+x').y = x.y + x'.y, (additivity).
4) x.x > 0 unless x = 0 (positivity, or positive definiteness).

In particular x = 0 iff x.y = 0 for all y. Recall that geometrically, x.y = 0 iff x and y are perpendicular (the zero vector is thus perpendicular to everything). We define the length |x| of x to be sqrt(x.x). Indeed the law of cosines says that x.y = |x||y|cos(t), where t is the angle in radians between the vectors x and y. This law gives the correction term to the Pythagorean theorem for non right triangles, and appears in Euclid in that form. So x is perpendicular to y if and only if cos(t) = 0, if and only if x.y = 0.

Complex dot products
Since R^n is contained in C^n, it is useful to extend the dot product to vectors in C^n. To maintain positivity, we need z.z to be real and positive, so if z = (z1,...,zn), and w = (w1,...,wn), we define z.w = (z1bar)w1+...+(znbar)wn, where zjbar is the complex conjugate of zj, i.e. if zj = aj+i bj, then zjbar = aj-i bj. The bar of a coordinate vector is obtained by taking the conjugate of every entry. then if [z],[w] are the column vectors represented by z,w, z.w = [zbar]^t [w]. The following properties hold for z, z',w in C^n, and c in C:
1) z.w = (w.z)bar, (conjugate symmetry).
2) (cz).w = cbar (z.w), (z.cw) = c(z.w), (sesquilinearity, i.e. "one and a half linearity").
3) (z+z'.w) = z.w + z'.w, (additivity).
4) z.z > 0 unless z = 0, (positive definiteness).

Again z = 0 iff z.w = 0 for all w. We do not define the angle between two complex vectors, but by analogy with the real case, we call z,w orthogonal iff z.w = 0. Similarly, we define length of z = sqrt(z.z) = |z|. Then length(z/|z|) = 1.

Definition: In general, any real or complex vector space equipped with a dot product satisfying these properties is called an inner product space.

Definitions: A basis is called orthogonal if any two distinct vectors in the basis are orthogonal, and is called orthonormal if orthogonal and all vectors have length one.

By the law of cosines, if y has length one, then x.y is the directed length of the projection of x onto the line spanned by y. Using this fact, we can replace the vectors in any sequence, by orthogonal vectors with the same span.
Theorem: A finite dimensional inner product space V always has an orthonormal basis.
proof: We give a recipe for reducing any basis to an orthonormal one. Suppose (x1,...,xn) is a basis, and set y1 = x1/|x1|, so that |y1|^2 = (y1.y1) = 1. Now consider x2, and project it onto Span(y1) by looking at (x2.y1)y1. Then note that x2 – (x2.y1)y1 is orthogonal to y1, since (x2 – (x2.y1)y1).y1 = (x2.y1) – (x2.y1)(y1.y1) = (x2.y1) – (x2.y1) = 0. Since y1 but not x2, is a multiple of x1, z2 = x2 – (x2.y1)y1 ≠ 0. Thus replacing x2 by z2/|z2| = y2, gives a pair (y1,y2) of orthonormal vectors with the same span as (x1,x2). Continue like this....Replace x3 by y3 = z3/|z3|, where z3 = x3 – (x3.y1)y1 – (x3.y2)y2. Then y3 has length one, is orthogonal to y1 and y2, and Span(y1,y2,y3) = Span(x1,x2,x3). Continuing like this produces an orthonormal basis for V. QED.

Remark: This application of the law of cosines, called the Gram Schmidt process, gives the following: for any finite sequence of vectors (x1,...,xn) in an inner product space, there is an orthogonal sequence (y1,...,yn) such that for each k, Span(x1,...,xk) = Span(y1,...,yk), and each yj has length 1 or zero. If xk belongs to Span(x1,...,xk-1) = Span(y1,...,yk-1), then xk = (xk.y1)y1 + ... +(xk.yk-1)yk-1, expresses xk as a linear combination of the orthogonal vectors (y1,...,yk-1). In this case zk = 0 = yk.

Definition: An isomorphism T of inner product spaces is called an isometry if it carries the inner product of one space into that of the other, i.e. if x.y = (Tx.Ty) for all x,y.

Cor: Every finite dimensional inner product space V is isometrically isomorphic to R^n or C^n, with the standard dot products.
proof: Find an o.n. basis for V. The resulting isomorphism with R^n or C^n, carries the inner product on V into the standard inner product on R^n or C^n. I.e. if (y1,...,yn) is an orthonormal basis for V, and the vectors x = a1y1+...+anyn, and x' = a1'y1+....+an'yn, correspond to a = (a1,...,an) and a' = (a1',...,an') in C^n, then since yj.yj = 1, and yi.yj = 0 when i≠j, by expanding we have x.x' = (a1y1+...+anyn).(a1'y1+...+an'yn) = a.a', equal to the dot product in C^n. **QED.**

The usefulness of orthogonality for us is in defining direct sum decompositions.
Define for any inner product space V and subspace, Uperp = {all vectors in V that are orthogonal to all vectors in U}. Then Uperp is a subspace complementary to U.

Lemma: If U is a subspace of V, then V is a direct sum of U and Uperp.
proof: There is no vector in both U and Uperp except zero, since zero is the only vector which is orthogonal to itself. By the construction above, if (x1,...,xn) is a basis for V such that (x1,....,xk) is a basis for U+Uperp, and if U+Uperp does not span V, then k < n, so xn is a vector orthogonal to U but not belonging to Uperp, a contradiction. **QED.**

Transposes and Adjoints.
Now we ask for links between a matrix A and dot products. Define A^t = transpose of A = the result of interchanging rows and columns of A. The basic property we need of this concept is that (AB)^t = B^t A^t. I.e. the entry in the ith row and jth column of AB is the dot product of the ith row of A with the jth column of B. Hence the ith row and jth column of (AB)^t is the dot product of the jth row of A and the ith column of B = dot product of ith row of B^t and jth column of A^t = the entry in the ith row and jth column of B^t A^t. Define the adjoint of A over C as Abar^t. Since (AB)bar = Abar Bbar, we get (AB)bar^t = (Bbar^t)(Abar^t).

Define the adjoint of A on R^n or C^n, as the linear map A* defined by multiplication by either the transpose in the real case, or the conjugate transpose in the complex case, of the matrix of A. The following property characterizes the adjoint.

Lemma: In both the real and complex cases, (Ax).y = x.(A* y).
proof: real case: (Ax).y = [Ax]^t [y] = x^t A^t y = x.(A^t y).
complex case: (Ax).y = [xbar]^t [Abar^t y] = [Abar xbar]^t [y] = [xbar]^t [Abar]^t [y] = [xbar]^t [Abar^t y] = x. (Abar^t y). **QED**.

Remark: If y,z are vectors in an inner product space V such that x.y = x.z for all x in V, then y = z, since then x.(y-z) = 0 for all x.

We can view the map given by a transpose or conjugate transpose intrinsically as well.

Cor: If T:V-->V is a linear map on any inner product space V, there is a unique linear map T* called the "adjoint of T" such that for all x,y in V, Tx.y = x.(T*y).
proof: For uniqueness, the previous remark implies for each y there is at most one choice of T*y, since any two choices have the same dot product with every x. For existence, since V is isometrically isomorphic with R^n or C^n, choose such an isomorphism, i.e. any orthonormal basis, and then take A* to be the map defined by the adjoint. **QED.**

Remark: T** = T. This is clear from the matrix definition, and abstractly, T*x.y = (y.T*x)bar = (Ty.x)bar = (x.Ty). Hence Ty satisfies the property needed to be T**y, so by uniqueness Ty is T**y.

Note: It follows that T is an isometry of V if and only if T* = T.

A more abstract argument for the corollary in the complex case is this. Let T:V-->V be given. Then Hombar(V,C) = {the space of conjugate linear maps V-->C}, is a complex vector space of dimension equal to dim(V). Moreover the map f:V-->Hombar(V,C) taking z to (.z) is an isomorphism, as is the map g:V-->Hombar(V,C) taking y to (T( ).y). Thus the composition (f^-1)o(g):V-->V takes y to the unique vector z such that for all x, we have (Tx.y) = (x.z). Thus we can define this z to be T*y, and this composition to be T*. In the real case, use the intermediate space Hom(V,R) = {linear maps V-->R}. This argument proves existence and uniqueness simultaneously, since the map f is surjective (existence) if and only if it is injective (uniqueness).

Maps that commute with their adjoints turn out to be especially understandable.
Definition: If V is an inner product space, T:V-->V is **normal** if TT* = T*T.

The structure of a normal linear operator is closely linked with that of its adjoint.
Proposition: If T:V-->V is normal, on a finite dimensional inner product space V, then
1) T and T* take each vector x to vectors of the same length, i.e. for all x, |Tx| = |T*x|.
2) T and T* have the same kernel, i.e. ker(T) = ker(T*).
3) T and T* have the same eigenvectors, with conjugate eigenvalues, i.e. if Tx = cx, then T*x = cbarx.
4) Eigenvectors of T with different eigenvectors, are orthogonal, i.e. if Tx = cx, and Ty = dy, with c ≠ d, then x.y = 0.
proof: 1) |Tx|^2 = (Tx.Tx) = (x.T*Tx) = (x.TT*x) = (T*x.T*x) = |T*x|^2.
2) Tx = 0, implies 0 = |Tx| = |T*x|, so T*x = 0.
3) Since (T-c)* = (T*-cbar), if x is in ker(T-c), then x is also in ker(T*-cbar).
4) If Tx=cx and Ty = dy with c ≠ d, then cbar(x.y) = (Tx).y = x.(T*y) = x.(dbary) = dbar(x.y). Since c ≠ d, also cbar ≠ dbar, so x.y = 0. **QED.**

Next we prove several results describing the structure of normal operators. Again there are cases which are easier than the general one, and it is useful to prove those first.
Lemma: For any operator T on an inner product space, if U is a T invariant subspace, then Uperp is T* invariant, and vice versa.
proof: easy exercise.

Definition: T is called hermitian, if T* = T.
Prop: If T is hermitian on a finite dimensional V, then T is orthogonally diagonalizable.
proof: Since the characteristic polynomial ch(t) has a linear factor, there is an eigenvector x. Then if x.y = 0, we have x.Ty = Tx.y = cx.y = 0, so (x)perp is also T invariant. T is still normal on (x)perp, so by induction on dimV, T is orthogonally diagonalizable on (x)perp, hence on all of V. **QED.**

Remark: If T is hermitian, then all eigenvalues of T are real.
proof: If Tx = cx, and x ≠ 0, then we have cbar(x.x) = (cx).x = Tx.x = x.T*x = x.Tx = x.cx = c(x.x). Since x ≠ 0, x.x ≠ 0, so c = cbar. **QED.**

The next case is only slightly more work.
Definition: T is called unitary, if T* = T^-1.

Prop: If dimV is finite and T:V-->V is unitary, T is orthogonally diagonalizable.
proof: Again ch(t) has a linear factor factor, so there is an eigenvector x. If T(x) = cx, then T^-1(x) = c^-1 x. So if x.y = 0, then x.Ty = T*x.y = T^-1x.y = c^-1 (x.y) = 0, so (x)perp is T invariant. Since T is normal on (x)perp, induction on dimV shows T is orthogonally diagonalizable on Uperp, hence on all of V. **QED.**

Cor: If T is unitary, all eigenvalues of T have absolute value one.
proof: T is unitary, so Tx.Ty = x.T*Ty = x.T^-1Ty = x.y. So if Tx = cx, then x.x = Tx.Tx = (cx).(cx) = cbar c (x.x). Hence cbar c = 1, i.e. |c| = 1. **QED.**

Recall T is normal if TT* = T*T. We give a slightly different inductive proof next.
Prop: If T is normal on V, and dimV finite, T is orthogonally diagonalizable.
proof: Since ch(t) has an eigenvalue c, there is a non zero eigenspace U = {x in V: Tx = cx}. Any basis for U consists of eigenvectors for T. If x is in U and y is in Uperp, then x.y = 0, and x.Ty = T*x.y. We claim T*x is in U. I.e. TT*x = T*Tx, since x is normal, and since Tx = cx, we have T*Tx = T*cx = cT*x, so T* belongs to U. Hence x.Ty = T*x.y = 0, since y is in Uperp. Thus Uperp is T invariant, T is normal on Uperp, and by induction on dimV, T is orthogonally diagonalizable on Uperp, hence on all of V. **QED.**

Next we take up the real cases.
Definition: A real operator on an inner product space is called symmetric if T = T*.
Prop: If T is (real) symmetric on V, dimV finite, then T is orthogonally diagonalizable.
proof: Choose a matrix A for T in some orthonormal matrix. Consider A as a matrix on C^n instead of R^n. Then A is hermitian, hence has all real eigenvalues. Since the eigenvalues are the roots of the characteristic polynomial of A, they are the same whether we regard A as a real or a complex matrix. Now that A has a real eigenvalue, it has a real eigenvector x. Then if x.y = 0, we have x.Ty = Tx.y = cx.y = 0, so (x)perp is T invariant. T is normal on (x)perp, so by induction on dimV, T is orthogonally diagonalizable on (x)perp, hence on V. **QED.**

Those four cases are the easiest. The fifth case is only a little more work.
Definition: A real operator T on V with dimV finite, is called orthogonal if T* = T^-1.

Prop: If T is real orthogonal on V, dimV finite/R, then V is an orthogonal direct sum of invariant subspaces of (real) dimension ≤ 2, on each of which T = ±Id, or a reflection.
proof: The first step is to produce an indecomposable subspace U of dimension ≤ 2.

Lemma: For any operator T on a finite dimensional space V, if the minimal polynomial has a non constant factor f, then there is a non zero vector x annihilated by f(T).
proof: If the minimal polynomial m = fg, and f is injective, then f(T)g(T)(x) = 0 for all x, implies that g(T)(x) = 0 for all x. Then g would be a polynomial of lower degree than m that annihilates all of V, a contradiction. **QED.**

Cor: If f is an irreducible factor of the minimal polynomial of T, there is an indcomposable invariant subspace U of V of dimension equal to deg(f).
proof: We know the cyclic subspace generated by x is indecomposable if f(T)x = 0 where f is irreducible. **QED.**

Hence there is an indecomposable invariant subspace U of dimension ≤ 2, corresponding to a real irreducible linear or quadratic factor of the minimal polynomial. Since T is unitary as a complex matrix, the complex and real eigenvalues of T have absolute value one, so the real eigenvalues equal ± 1. Thus on every one dimensional indecomposable invariant subspace, T = ±Id.

If U has dimension 2, since T is length and angle preserving on U, if (x1,x2) is an orthonormal basis for U, Tx1 is length one and orthogonal to Tx2. Hence T(x1) = cos(a)x1 + sin(a)x2, and Tx2 = -sin(a)x1 + cos(a)x2, or else Tx2 = sin(a)x1 - cos(a)x2.

But the minimal polynomial on U is irreducible over R. If T(x1) = cos(a)x1 + sin(a)x2, and Tx2 = Tx2 = sin(a)x1 - cos(a)x2, the minimal polynomial equals t^2 – 1, which is reducible. So in fact, T(x1) = cos(a)x1 + sin(a)x2, and Tx2 = -sin(a)x1 + cos(a)x2, which is a rotation. Since T is normal on Uperp, which is T invariant, by induction on dimV the theorem is proved. **QED.**

The final version for real normal operators takes more effort, but is not too bad.
Prop: If V is a real space, and T:V-->V a normal operator with minimal polynomial m(t) = ∏(t-ci)^di∏qj^ej, all ci distinct real scalars, and all qj distinct irreducible real monic quadratic polynomials, then:
i) V is an orthogonal direct sum of the invariant subspaces ker∏(T-ci) and ker∏qj(T).
ii) The eigenspaces ker∏(T-ci) decompose into one dimensional invariant subspaces, and the invariant subspaces ker∏qj(T) decompose into indecomposable invariant two dimensional subspaces, on each of which qj is the minimal polynomial.
iii) In particular all di =1, and all qj =1. The matrix of T on the eigenspace ker∏(T-ci) is diagonal with ci along the diagonal, and the matrix of T on ker∏qj(T) is a block matrix with 2 by 2 matrices of form
|aj -bj |
|bj aj |, along the diagonal, where the roots of qj are aj ± i bj.
proof: Again we have an indecomposable invariant subspace U of dim ≤ 2. If dimU = one, it is an eigenspace. If dimU = 2, we need to discuss T restricted to U. We claim that

in any orthonormal basis for a normal T on a two dimensional invariant subspace U where the minimal polynomial is (t-c)(t-cbar), with c = a+bi, that T has matrix
| a -b |
| b a |, or its transpose, if we interchange c and cbar.

To see this, let the matrix be
| x y |
| z w |.

Since T is normal the lengths of the first column and row are equal, so z^2 = y^2, so -z = y, or z = y. But the characteristic polynomial is f = t^2 – (x+w)t +xw –yz = (t-c)(t-cbar), with c = a+bi, i.e. f = t^2 -2at + (a^2+b^2). Since this polynomial is irreducible over R, its discriminant is negative, i.e. 4a^2 < 4(xw-yz). But since x+w = 2a, we have xw ≤ a^2. So 4a^2 < 4(xw-yz) ≤ 4a^2 -4yz. This implies –yz > 0, so y and z have opposite sign, hence y = -z.

Then since this matrix commutes with its transpose, we get that x = w. Since then the characteristic polynomial is t^2 – 2xt + (x^2+z^2), when it should be t^2 - 2at+(a^2+b^2), it follows that a = x = w, and b = z = -y, or b = y = -z.

Next we show the orthocomplement of the invariant U just produced is also T invariant.
Note: Recall that in any orthonormal basis, the columns of the matrix of T* are the rows of the matrix of T. Thus property 1) implies that if T is normal, the length of the jth column of T equals the length of the jth column of T*, i.e. the length of the jth row of T. I.e. for the matrix of a normal T, the sum of the squared absolute values of the jth row, equals the sum of the squared absolute values of the jth column. This is the key to the next important result.

Orthogonal decomposition lemma: If T:V-->V is normal, and U is a T invariant subspace, then so is Uperp.
proof: Choose an orthonormal basis (x1,...,xn) for U and an orthonormal basis (y1,...,ym) for Uperp. Since U is T invariant, the matrix for T in this basis looks like this:
| A B |
| 0 C |,
i.e. [T] has a block of zeroes in the lower left corner, because A is the n by n matrix of T restricted to U, and for all i, T(xi) only depends on the basis (x1,...,xn) of U. We claim the block matrix B is also zero, which will imply that for all j, T(yj) depends only on the basis (y1,...,ym) of Uperp, i.e. T(Uperp) is a subspace of Uperp, as desired. Since the basis is orthonormal, the sum of the squares of the absolute values of the entries in the jth column of A equals the squared length of Txj. For the same reason the sum of the squares of the absolute values of the entries in the ith row of | A B | equals the squared length of the vector T*xi, which has the same length as Txi. Since the sum of the squares of the lengths of the vectors (Tx1,...,Txn) equals the sum of the squares of the lengths of the vectors (T*x1,...,T*xn), the sum of the squares of the absolute values of the entries of A, equals the sum of the squares of the entries of | A B |. Hence the squares of the absolute values of all entries in B equals zero, so B = [0]. This proves the lemma, and allows us again to finish the proof by induction on dimV. **QED.**

So in some orthonormal basis, the matrix of a real normal operator looks like this:

```
| a1  -b1 |  0...........                              ....0|
| b1   a1 |  0............                            .....0|
| 0   0 | a2   -b2| 0.......                          .....0|
| 0   0 | b2    a2| 0.........                         ....0|
......................
| 0   0  0    ...0 | an   -bn | 0......               .....0|
| 0   0  0    ...0 | bn    an | 0.....                .....0|
..........

| 0...........                  ........0 |c1| 0.......... 0|
| 0 ........................................0  0  | c2| 0......0|
| 0 ..........

| 0 ........................................................0 |cm|,
```

where not all the a's, b's, and c's need be different.

Remark: Instead of separating the proofs by difficulty, we could have combined them.
Outline of proofs of spectral thms:
Step one: For any operator T on a finite dimensional space V, if the minimal polynomial has a non constant factor f, then there is a non zero vector x annihilated by f(T).
proof: If the minimal polynomial m = fg, and f is injective, then f(T)g(T)(x) = 0 for all x, implies that g(T)(x) = 0 for all x. Then g would be a polynomial of lower degree than m that annihilates all of V, a contradiction. **QED.**

Cor: If f is an irreducible factor of degree d in the minimal polynomial of T, there is an indecomposable invariant subspace U of dimension d in V, with f = the minimal polynomial of T restricted to U.
proof: Let U = the T cyclic subspace spanned by a non zero x such that f(T)(x) = 0. Then (x,Tx,...,T^(d-1)x) spans U and any lower degree polynomial in T annihilating x would divide f, a contradiction to irreducibility of f. **QED.**

Step two: If T is normal on V, any invariant subspace of V has an invariant orthocomplement.
proof: This is proved by the orthogonal decomposition lemma. **QED.**

Cor: Structure of normal operators.
Assume T:V-->V is a normal operator on a finite dimensional inner product space.
1) If V is a complex space, with minimal polynomial m(T) = ∏(t-cj)^dj, all cj distinct, then V decomposes into an orthogonal direct sum of eigenspaces Vj = ker(T-cj). I.e., all dj = 1, and there is an orthonormal basis of V in which the matrix of T is diagonal.

2) If V is a real space, with minimal polynomial m = ∏(t-ci)^di∏qj^ej, all ci distinct real scalars, and all qj distinct irreducible real monic quadratic polynomials, then
i) V is an orthogonal direct sum of the invariant subspaces ker∏(T-ci) and ker∏qj(T).

ii) The eigenspaces ker∏(T-ci) decompose into one dimensional invariant subspaces, and the invariant subspaces ker∏qj(T) decompose into indecomposable invariant two dimensional subspaces, on each of which qj is the minimal polynomial.
iii) In particular all di =1, and all qj =1. The matrix of T on the eigenspace ker∏(T-ci) is diagonal with ci along the diagonal, and the matrix of T on ker∏qj(T) is a block matrix with 2 by 2 matrices of form
|aj -bj |
|bj aj |, along the diagonal, where the roots of qj are aj ± i bj.
We get all the theorems from steps 1) and 2) by induction.

You may find the following lemma useful.
Lemma: If A is a real n by n matrix, it is also a complex n by n matrix, and it has the same minimal polynomial over C as over R.
proof: Certainly, the minimal polynomial over C divides the one over R, so it remains to show they have the same degree. We claim that any sequence of vectors in R^n which is independent over R, remains independent over C. To see this recall that we can put these real vectors into the columns of a matrix and reuce the matrix over R. Then the independence is revealed in the fact that all the columns become pivot columns after reduction. But now even over C, pivot columns are always independent, i.e. they are also pivot columns viewed over C, so no non trivial C linear combination of pivot columns can equal zero. Now view the space Mat,nxn(R) as R^N, where N = n^2. Then it is a subspace of Mat,nxn(C) = C^N. So a sequence of real matrices is independent in Mat,nxn(R) if and only if it is independent in Mat,nxn(C). Thus if the real minimal polynomial of T has degree ≥ n, then Id,T,T^2,....,T^(n-1) are independent over R, hence also over C. Thus the minimal polynomial of T over C also has degree ≥ n. Since the degree over C is ≤ the degree over R, the two polynomials are both monic of the same degree, hence equal. **QED.**

Remark: This implies that a real matrix which is diagonalizable as a complex matrix, must have minimal polynomial consisting of distinct irreducible factors in R[t].

Terminology: The results usually called “spectral theorems” are those whose conclusion is that the matrix is diagonalizable, three complex ones and one real symmetric one.

Orthogonality has an abstract version, without inner products.

Dual spaces, dual bases and adjoints

We have often discussed the columns of a matrix, but have not said much about the rows. I.e. given a linear map T:V-->W, and a basis (x1,...,xn) for V, we get a matrix of “columns”, where the jth column is [T(xj)]. If we also have a basis (y1,...,ym) for W, this “column” becomes an actual column of scalars, namely the coordinates of T(xj) in the basis (y1,...,ym) for W. Consequently, in the matrix of T, the ith row consists of the sequence (Ti(x1),...,Ti(xn)) of ith coordinates of the column vectors of [T].

So just as a map T:k^n-->W is determined by its sequence (T(x1),...,T(xn)) of values on the standard basis of k^n, so also is a linear map T:V-->k^m determined by its sequence (T1,....,Tm) of compositions Ti:V-->k, of T with the m projections onto coordinates of vectors in k^m. Thus while the columns of [T] represent vectors in W, the rows represent functions V-->k.

Looked at another way, we know a linear map ∏Vj-->W is determined by its restrictions to each factor Vj-->W. Hence Hom(k^n,W) = ∏Hom(k,W), where if T:k^n-->W is a map, then its components in the factors of the product ∏Hom(k,W) are represented by the columns of [T]. To see again that columns are vectors in W, we claim Hom(k,W) = W. I.e. the natural map Hom(k,W)-->W sending f to f(1), is an isomorphism. It is injective since f(1) = 0 implies f(t) = 0 for all t in k, so f = 0. It is surjective since for any y in W, the map f(t) = tw, has f(1) = y.

But rows behave differently. A linear map T:V-->∏Wi is determined by its compositions V-->Wi to the factors of W. I.e. Hom(∏Vj, W) = ∏Hom(Vj,W), and Hom(V,∏Wi) = ∏Hom(V,Wi). Thus Hom(V,∏k^m) = ∏Hom(V,k), and each map T: V-->∏k^m decomposes into components Ti in Hom(V,k) corresponding to rows of the matrix [T], but there is no natural isomorphism from Hom(V,k) to V. So a row of [T] representing an element of Hom(V,k), is not naturally an element of V, but should be considered a linear function V-->k. So we should distinguish the following concept.

Definition: Given a vector space V, define the “dual space of V” = Hom(V,k) = V*.

I admit it is not easy to tell the difference between a row vector of length n and a column vector of length n, i.e. there is little visible difference between an element of Hom(k^n,k) and an element of k^n. But this only because here we have a chosen finite basis. I.e. if we are given a finite basis of V, there is an identification between V and V*.

Dual bases

If (x1,..,xn) is a basis of V, define elements of V* as follows: set fj = the unique linear function V-->k such that fj(xi) = 0 for i ≠ j, and fj(xj) = 1. Then (f1,...,fn) is a basis for V*, called the basis dual to (x1,...,xn). To see it is a basis, let f:V-->k be any linear function, and note that f = f(x1)f1+...+f(xn)fn, since both sides of this equation agree on all the basis vectors xj. Thus the sequence (f1,...,fn) spans V*. And if f = c1f1+...+cnfn = 0, then evaluating f on (xj), we see that 0 = f(xj) = cj, so all cj = 0, hence the sequence (f1,...,fn) is independent. Given a basis (x1,...,xn) for V, we thus get a dual basis (f1,...,fn) for V*, and hence an isomorphism V-->V* sending xj to fj, but this isomorphism changes when we change the basis of V. E.g. replacing x1 by 2x1 changes f1 into f1/2, changing the isomorphism.

Remark: It is important to notice that these results are true only in finite dimensions. If V has countable dimension, with basis (x1,....,xn,.....) and if (f1,...,fn,.....) is defined as above, then a (necessarily finite) linear combination of the fj, can only be non zero on a finite number of the basis vectors xj. But there is a linear function f that equals 1 on every xj. Hence this f cannot be in the span of the sequence (f1,...,fn,....). Indeed, if k = Z/2, then V has countable dimension, but V* has uncountable dimension, since a linear function on V is give by an arbitrary sequence of elements of Z/2. In particular V is not isomorphic to V*. On the other hand, the map Hom(k,W)-->W is defined and an isomorphism for every W. We will assume our vector spaces are finite dimensional from now on, unless we say otherwise.

Natural pairings
While V and V* are not naturally isomorphic, there is a natural pairing VxV*-->k, defined simply by evaluating a function on its argument, i.e. <x,f> = f(x). This pairing is non degenerate in the sense that for each f ≠ 0, there is by definition some x such that f(x) ≠ 0, and for each x ≠ 0 there is (by the dual basis construction) some f such that f(x) ≠ 0.

Thus an isomorphism f:V-->V* gives a non degenerate pairing VxV-->k, defined by setting <x,y> = f(y)(x). Conversely, a non degenerate pairing VxV->k defines an isomorphism V-->V*. Thus we can imitate some concepts arising from an inner product on V such as orthogonality, if we formulate it as a relation between vectors in V and V*.

Definition: Vectors x in V and f in V* are called orthogonal if f(x) = 0. If U is a subspace of V, Uperp is the subspace of all f in V* such that U is contained in ker(f).

Proposition: If U is a subspace of V, the natural map from (V/U)*-->V*, sending a linear function f:V/U-->k to the composition V-->V/U-->k, defines an isomorphism (V/U)*-->Uperp.

Proof: We know a linear map f:V-->k factors through a linear map V/U—k if and only if U is contained in ker(f). I.e. if it factors then U certainly goes to zero. And if f annihilates U, then it defines a function on elements of V/U. (Exercise) **QED.**

Cor: For any subspace U of V, dim(U) + dim(Uperp) = dim(V).
proof: dim(Uperp) = dim(V/U)* = dim(V/U) = dim(V) – dim(U). **QED.**

Cor: If V is a direct sum of U1 and U2, then V* is the direct sum of U1perp and U2perp.
proof: U1perp and U2perp are subspaces of V*, and any f in both of them annihilates U1 and U2, hence all of V, so is zero. But by the previous corollary, the sum of the dimensions of U1perp and U2perp equals dim(V), so they span it. **QED.**

Cor: If V is a direct sum of U1 and U2, then V* = (U1+U2)* is isomorphic to the direct sum of (U1)* and (U2)*.
proof: This is a special case of the principle that a linear map U1xU2-->k is determined by its restrictions to U1 and U2, and those restrictions can be chosen arbitrarily, i.e. for any spaces U,V,W, Hom(UxV,W) = Hom(U,W) x Hom(V,W). And we know the direct sum of two subspaces is isomorphic to their direct product. **QED.**

Some statements about orthogonal complements seem not to make sense in this setting. E.g. for any subspace U in an inner product space V, we have Uperp,perp = U, but here Uperp,perp is a subspace of V**, not V. Fortunately V and V** are essentially the same in finite dimensions.

Prop: For any vector space V, there is a natural linear injection of V into V**, sending a vector x from V, to the function ev(x) = evaluation at x, on V*. Hence in finite dimensions, where V and V** have the same dimension, this is an isomorphism.

proof: We map x to ev(x), a function V*-->k, defined as follows. If f:V-->k is an element of V*, then ev(x)(f) = f(x). Since if x ≠ 0, we can always choose a basis for V which ahs x as one element, there is always an f in V* that sends x to 1, so ev(x)(f) = 1 for that f, hence if x ≠ 0 then ev(x) ≠ 0. I.e. the map V-->V** is injective. It is linear because functions in V* are linear. **QED.**

Exercise: If V has finite dimensions, the isomorphism V-->V** carries each subspace U of V isomorphically onto Uperp,perp in V**. **[**Hint: Show U maps into Uperp,perp, and the two subspaces have the same dimension.]

Abstract adjoints

It is a general principle in modern mathematics that maps are even more important than spaces, indeed maps are the best way to compare spaces, and any construction that makes new spaces out of old, should also make new maps out of old ones. Next we show that the construction of the space V* out of V, leads naturally to construction of a map T* out of T. The definition makes sense in infinite dimensions.

Definition of T* without an inner product

Let T:V-->W be any given linear map, and define T*:W*-->V* to be "preceding by T". I.e. for f:W-->k in W*, define T*f = (foT):V-->W-->k in V*. Then we have two basic properties, (IdV)* = Id(V*) and for any two maps S,T, with T:V-->W, and S:U-->V, so that (TS):U-->W, we have (TS)* = S*T*:W*-->U*. This follows immediately from associativity of composition. I.e. (TS)*(f) = f(TS) = (fT)S = (T*f)S = S*(T*f) = **S***T*f.

The next property of an adjoint is fundamental.

Proposition: For any map T, we have ker(T*) = (ImT)perp.

proof: Since T*(f) = 0 if and only if the composition fT = 0, this means precisely that f is zero on the image of T. **QED.**

Note the previous proposition also holds without any finite dimensional assumptions. The analogous claim that ker(T) = (ImT*)perp however faces the problem that ker(T) is a subspace of V while (ImT*)perp is a subspace of V**. Hence to deduce it from the previous proposition would seem to require finite dimensionality to identify V with V**. The following version however is true without that assumption.

Proposition: For any T:V-->W, we have (kerT)perp = Im(T*), as subspaces of V*.

Proof: The image of T* equals those functions in V* which are compositions of form (fT):V-->W-->k. But a function g:V-->k factors through the map T:V-->W if and only if it annihilates kerT. I.e. if g annihilates kerT, then g induces a linear map V/kerT-->k. But the map T:V-->W induces an embedding of V/kerT into W. So it suffices to define f:W-->k by extending g from the subspace V/kerT to all of W. This is done by extending a basis of V/kerT to a basis of W and setting f = 0 on the remaining vectors. The other direction is trivial since if T annihilates x, so does fT, for all f. **QED.**

The reader is welcome to consider only finite dimensional spaces, but we wanted to indicate to some extent where the hypothesis is used, and where it is not.

The double dual of a map
Having shown that V is naturally a subspace of V**, and essentially equal to it in finite dimensions, it should follow that T** is a natural extension of T, and essentially equal to it in finite dimensions. We check that next.

Proposition: If T:V-->W is any linear map, then natural embeding V-->V** makes T correspond to T**. I.e. if x in V and y in W correspond to x** in V** and y** in W**, then [T(x)]** = T**(x**).
proof: [T(x)]** is an element of W**, such that if g is in W* then [T(x)]**(g) = g(T(x)). But T**(x**) is the element such that for g in W**, T**(x**)(g) = x**(T*(g)) = x**(gT) = gT(x). These are the same. **QED.**

Finally, the abstract adjoint should correspond to the transpose in terms of appropriate bases. Indeed it does.
Proposition: If (x1,...,xn) is a basis for V, and (y1,...,ym) a basis for W, and T:V-->W a linear map, then in terms of these bases and theor dual bases for V* and W*, the matrix of T* is the transpose of the matrix for T.
proof: We denote the dual bases of V* and W*, by (x*1,...,x*n) and (y*1,...,y*m). Then the (i,j) entry of the matrix of T, is the ith coordinate in the basis (y1,...,ym), of the element T(xj), i.e. it equals y*i(Txj). Now the (j,i) entry of the matrix for T* is the jth cordinate in the basis (x*1,...,x*n), of the element T*(y*i), i.e. it is x**j(T*(y*i)) = T*(y*i)(xj) = y*i(Txj). Since this equals the (i,j) entry of [T], the two matrices are transposes of one another. **QED.**

Remark on adjoints in analysis: The isomorphism from a space V to its dual V* in finite dimensions gives a way to describe a vector in V by saying only how it behaves as a function on V, i.e. by describing the corresponding function in V*. E.g. we defined the element T*y by saying it equals the unique vector in V that defines the same function of x as Tx.y. I.e. we did not define the element T*y, we defined the function f(x) = Tx.y, and then said this function must equal x.T*y for some unique element T*y.

This trick works wonders in analysis where we are dealing with infinite dimensional spaces which are not always equal to their duals. Then we are concerned with identifying which functionals on a given do in fact come from V. E.g. consider the space S of smooth functions on R and the subspace Sc of compactly supported smooth functions. Differentiation is defined as an operator D on these spaces. A larger space I of locally integrable functions can be viewed as a subspace of the dual Sc* of Sc, by letting a locally integrable function f act on a smooth function with compact support by integrating their product. Thus I, and S, embed as subspaces of the dual Sc* of the smaller space Sc. We would like to extend the differentiation operator D to the space I, but how are we to differentiate functions that are only integrable and not differentiable? Recall that an integrable function can fail to be differentiable at many points, and even to have many discontinuities.

Since I is embedded in the space Sc*, it suffices to tell how to define the differentiation operator on Sc*, i.e. to tell how to differentiate a linear functional on Sc. So what we will do is define the derivative on the subspace I of S*, but with values that belong to Sc*. I.e. we will define the derivative as a map D:I-->Sc*, as follows. If we are given a compactly supported smooth function g in Sc, and a locally integrable function f, we want to define how Df acts on g. By the formula for integration by parts, since the product fg vanishes at infinity, the integral of Df.g + fDg should be zero, so we should have the integral of Df.g equal to minus the integral of f.Dg. So we define <Df,g> = integral of -f.Dg. This makes perfect sense because g is smooth. More generally, if L is any linear function in Sc*, define DL as a linear function on a smooth g to be L(-Dg). I.e. D:Sc*-->Sc* is defined as the adjoint of the map (-D):Sc-->Sc, so DL(g) = L(-Dg).

Then we can ask for solutions in Sc* of a differential equation like Dy = L, where L is a linear function in Sc*. I believe it turns out this always has a solution where y is a locally integrable function. Moreover I believe there is a regularity lemma that a solution y of Dy = g, is smooth whenever g is smooth. This technique is used to solve other important equations, like the famous dbar equation dbar(y) = g in complex analysis. I.e. when g is a smooth function, one proves there is always an element L in Sc* such that dbar(L) = g, and then afterwards, that since g is smooth, in fact L is represented by a smooth function f. The reason this approach is natural is that one looks for solutions of differential equations by successive approximations, finding a sequence that converges to a solution. So one needs a complete space in which good sequences do converge. Since S is not complete in the integral norm appropriate to the problem, it helps to embed it in the larger space Sc* that is more likely to be complete. Then the approximate solutions do converge to a solution in Sc*. Finally one wants the solution to be a smooth function, so after getting it as a linear functional, one then proves it is in fact represented by a smooth function. This beautiful technique is called the theory of distributions.

The characteristic polynomial of a linear map

Everything proven so far has been dependent on knowing the minimal polynomial of our map T. We have given an algorithm for computing minimal polynomials, but it is useful also to understand their relation to the "characteristic polynomial" defined via determinants. We assume the reader knows determinants, which are treated afterwards.

Definition: If A is an n by n matrix the characteristic polynomial ch(t) = det[tId-A].

By the multiplicative property of determinants, A and Q^-1AQ have the same characteristic polynomial, hence every map T:V-->V on a finite dimensional space has a unique characteristic polynomial, defined in terms of any convenient basis.

Corollary: (Cayley Hamilton) If T:V-->V is a linear map on a finite dimensional space, then ch(T) = 0, i.e. T satisfies its own characteristic polynomial. Equivalently, if m(t) is the minimal polynomial, m(t) divides ch(t), in k[t].

proof: It suffices to prove it for n by n matrices A over k. If the minimal polynomial factors into linear factors over k, this is visible from the Jordan form, since if m = ∏(t-ci)^ni, there is a block in the Jordan form whose characteristic polynomial is (t-ci)^ni, and this divides the characteristic polynomial of A. I.e., a Jordan matrix J is in upper diagonal form, as is (tId-J), hence the determinant of (tId-J) is just the product of

the diagonal entries, i.e. of the various factors (t-ci) on the diagonal, and for each i, there are at least ni factors of (t-ci).
In the general case, extend the base field k to a splitting field for ch. The theorem holds for a Jordan matrix J = Q^-1 A Q, over that field. I.e. ch(J) = 0, where J and A are similar hence have the same characteristic polynomial. But then A = QJQ^-1, so for every n, A^n = (QJQ^-1)^n = Q(J^n)Q^-1, so ch(A) = Q(ch(J))Q^-1 = Q[0]Q^-1 = [0].
QED.

Remark: We will give a proof of Cayley Hamilton, using just LaGrange's inductive formula for determinants, after we discuss determinants and Cramer's rule.

It is clear from the Jordan form that the minimal polynomial and the characteristic polynomial have the same roots in k, but this can be proved directly.
Lemma: If m(t) and ch(t) are the minimal and characteristic polynomials of T, then m(a) = 0 for a in k, if and only if ch(a) = 0.
proof: If m(a) = 0, then m(t) = (t-a)f(t), for some polynomial f, by the famous root-factor theorem. Since m is the minimal polynomial for T, f(T) cannot annihilate all vectors in V, so (T-a) has a non zero kernel. Then the determinant of (T-a) is zero, i.e. t = a is a root of ch(t) = ± det(T-tId). Conversely if m(a) ≠ 0, then m = (t-a)f(t) + r, where r ≠ 0. Applying this to T, gives 0 = (T-a)f(T) + rId, hence rId = (T-a)(-f(T)), so (-1/r)f(T) is an inverse for (T-a). Since T-a has an inverse, it has no non trivial kernel. **QED.**

Thus if T has characteristic polynomial ch(t) = ∏(t-ci)^ni, then the minimal polynomial has form m(t) = ∏(t-ci)^di, where for each i, 1 ≤ di ≤ ni. The sum of the ni = dimV.

Determinants

To compute Jordan forms of a given map T, we need to know which scalars c have the property that T-c has a non trivial kernel. It is useful to have a formula in terms of entries of a matrix to determine if the matrix is invertible. One approach is to compute the oriented "volume" of the block spanned by the columns. I.e. this n - dimensional volume would be zero if and only if the columns are dependent. E.g. in a 2 by 2 matrix, the columns are dependent if and only if the parallelogram they span lies in a line, hence has zero area.

This suggests some properties such a formula should have. Since the volume of a block scales by c when we multiply one of the vectors spanning it by c, our volume function ∂(x1,...,xn) should satisfy: ∂(x1, ...,cxi,...,xn) = c ∂ (x1, ...,xi,...,xn). And when we stack two blocks on top of one another by adding two vectors in one entry, the volumes should add. So we should have ∂(x1,...,xi+yi,...,xn) = ∂(x1,...,xi,...,xn) + ∂(x1,...,yi,...,xn). Moreover, when two entries are equal, the vectors spanning the block are dependent, and the block lies in a lower dimensional space, so we should get zero: ∂(x1,...,xi,...,xi,...,xn) = 0. Finally the volume of the unit block spanned by the standard unit vectors should be 1.

So we want a function of n vector variables ∂:(k^n)^n --->k, that has these properties:
i) "alternating": ∂ equals zero when two entries are equal,
ii) "multilinear": ∂ is linear in one variable at a time, and

iii) “normalized”: on the block spanned by the standard unit vectors, ∂ has value 1.

Definition: An n dimensional “determinant” is a function ∂ on n by n matrices such that $\partial(Id) = 1$, and ∂ is alternating, and multilinear as a function say of the columns.

We will show determinant functions exist and are unique, and describe their properties.
Lemma: A determinant function is skew symmetric, i.e. ∂ changes sign when any two entries are exchanged: $\partial(x1,...,xi,...,xj,...,xn) = - \partial(x1,...,xj,...,xi,...,,xn)$.
proof: Since ∂ is alternating, $\partial(x1,...,xi+xj,...,xi+xj,...,xn) = 0$, and since it is multilinear, this equals $0 = \partial(x1,...,xi,...,xi,...,xn) + \partial(x1,...,xj,...,xj,...,xn) + \partial(x1,...,xi,...,xj,...,xn) + \partial(x1,...,xj,...,xi,...,xn)$. Since the first two entries are zero, we have our result. **QED.**

Remark: Conversely, skew- symmetric functions are almost alternating, since if $\partial(x1,...,xi,...,xj,...,xn) = - \partial(x1,...,xj,...,xi,...,,xn)$, then $2\partial(x1,...,xi,...,xj,...,xn) = 0$. This implies $\partial(x1,...,xi,...,xj,...,xn) = 0$, except in characteristic 2. But over the real field e.g., there is no difference in the two properties, at least for multilinear functions.

We will appeal to a few facts about permutations which we recall. In particular, a permutation of the integers $\{1,...,n\}$ is a bijective function $s:\{1,...,n\}\text{-->}\{1,...,n\}$.
Cor: Permuting the entries of ∂ by a permutation multiplies the value of ∂ by the sign of the permutation.
proof: Recall that the sign of a permutation which exchanges two entries is -1, and the sign of a composition of permutations is the product of their signs. One proof of this is to consider the action of the group of permutations on the ring of polynomials in n variables, $Z[X1,...,Xn]$ by permuting the variables. This action takes the polynomial $\prod i>j\ (Xi-Xj)$ into itself or minus itself, and we define the sign of the permutation accordingly. The permutation that exchanges X1 and X2, changes the sign of this polynomial since only the sign of $(X2-X1)$ is changed. Similarly every exchange of two adjacent variables changes the sign, and it can be deduced that every exchange of any two variables changes the sign. Since every permutation is a composition of such exchanges, the sign of every composition of permutations is the product of their signs. **QED.**

Uniqueness of determinants

Since a determinant ∂ is multilinear, we regard it as a kind of multiplication. Writing $xj = a1j\ e1 +.....+anj\ en$, in terms of the standard basis, and using multilinearity to expand our product $\partial\ (x1, x2,..., xn)$, gives n^n terms. I.e. $\partial\ (x1, x2,...,xn)$ = sum over all functions $f:\{1,..,n\}\text{-->}\{1,..,n\}$ of $\partial(a(f(1),1)\ e(f(1)),..., a(f(n),n)\ e(f(n)))$ = sum over all f of $\prod j:\ a(f(j),j)\ \partial(e(f(1)),...,e(f(n)))$. Since $\partial = 0$ on any sequence with repeated entries, this is really a sum over all bijective functions f, i.e. all permutations $s:\{1,...,n\}\text{-->}\{1,...,n\}$.

But since interchanging two entries multiplies the value of ∂ by -1, changing the entries by a permutation s, multiplies the value by sgn(s) = sign of the permutation s. So $\partial(x1,...,xn)$ = Sum over s: $sgn(s) \prod j\ a(s(j),j)\ \partial(e(1),...,e(n))$. Thus the value of a multilinear alternating ∂ is completely determined on every sequence $(x1,...,xn)$, by the value of $\partial(e(1),...,e(n))$. And for the standard determinant we decreed this value is one. Thus there is at most one way to define a determinant satisfying our three properties, namely by the formula: $\partial(x1,...,xn)$ = Sum over all permutations s: $sgn(s) \prod j\ a(s(j),j)$.

Define: ∂:Mat nxn (k) --> k, by ∂ ([aij]) = Sum over s: sgn(s) ∏j: a(s(j),j), the only possible multilinear alternating function in the columns of each matrix with ∂(Id) = 1.

The next property is fundamental.
Lemma: ∂(AB) = ∂(A)∂(B), for two nxn matrices A,B.
Proof: Since BA is a linear function of the columns of A, and ∂ is alternating and multilinear in those columns, the composite function d(A) = ∂(BA) is also multilinear and alternating in the columns of A, hence by our uniqueness argument, it is a constant multiple of ∂(A), and the multiplier is d(Id) = ∂(B). Hence ∂(BA) = ∂(B)∂(A). **QED.**

Cor: If A is invertible, then det(A) ≠ 0.
proof: If AB = Id, then det(A)det(B) = det(Id) = 1. **QED.**

Existence of the determinant function.
We must show the function defined by the formula above does have our three properties. If [aij] = Id, all but one term in ∂[Id] is 0, and that one term = 1, so normalization is satisfied. It remains to show this definition of ∂ is indeed multilinear and alternating.

Scaling is the easiest.
Lemma: ∂(x1,...,cxi,...,xn) = c ∂(x1,...,xi,...,xn), for all c in k, all x1,...,xn in k^n.
proof: The definition of ∂ is an alternating sum of terms ∏j: a(s(j),j), and in each of these terms there is one factor a(s(i),i). In the definition of ∂(x1,...,cxi,...,xn), this one factor a(s(i),i) is replaced by c(a(s(i),i)). Thus the whole sum is multiplied by c. **QED.**

Additivity is easy as well.
Lemma: ∂(x1,...,xi+yi,...,xn) = ∂(x1,...,xi,...,xn) + ∂(x1,...,yi,...,xn).
proof: . For simplicity we assume i = 1. Again each term of the sum defining ∂, is a product (a(s(1),1)+b(s(1),1))∏j>1: a(s(j),j) , where y1 = the column [b11 b21......bn1]^t. Since multiplication is distributive, this product is additive in each factor, i.e. (a(s(1),1)+b(s(1),1))∏j>1: a(s(j),j) = a(s(1),1)∏j>1: a(s(j),j) + b(s(1),1))∏j>1: a(s(j),j). Since a sum of additive functions is also additive, we are done. **QED.**

The alternating property follows as well, by grouping terms of the sum.
Lemma: If A has two equal columns, then det(A) = 0.
proof: If t is the involution interchanging say columns 1 and 2, then ts acts the same as s on all other columns but interchanges the value of s at columns 1 and 2, so a(s(j),j) = a(st(j),j) for j > 2. But every entry of column 1 equals the corresponding entry of column 2, so a(s(1),1) = a(s(1),2) = a(st(2),2), and a(s(2),2) = a(st(1),2) = a(st(1),1). Thus the two products ∏j a(s(j),j) and ∏j a(st(j),j) are equal, but they occur in the sum defining det(A) with opposite signs since sgn(st) = sgn(t)sgn(s) = -1sgn(s). Thus every term in the sum for detA is canceled by another term which is equal but of opposite sign. detA = 0. **QED**

Hence ∂(A) is indeed alternating and multilinear in the columns of A. Thus there does exist one and only one function satisfying the properties of a determinant.

Lemma: If A^t = [aji] is the transpose of A = [aij], then ∂(A^t) = ∂(A).
proof: ∂(A^t) = the sum over all s, of sgn(s)∏j a(j,s(j)), and if we reorder the factors in this product by the second subscript, this equals sgn(s)∏j a(s^-1(j),j). Since s and s^-1 have the same sign, this is also sgn(s^-1)∏j a(s^-1(j),j). Summing over the inverse of all permutations is the same as summing over all permutations, so this equals the sum over all s, of sgn(s)∏j a(s(j),j) = ∂(A). **QED.**

It is often useful to simplify a matrix by row reduction to introduce more zeroes before computing a determinant. We already know how scaling and interchanging rows affect the determinant, so we observe the third row operation does not change the determinant. Since detA = detA^t, we have all the same properties for rows as for columns.

Lemma: Adding a scalar multiple of a row to another does not change the determinant.
proof: Suppose we add c times the jth row xj to the ith row xi. Then by additivity, ∂(x1,..,xi+cxj,..,xj,..,xn) = ∂(x1,..,xi,..,xj,..,xn) + ∂(x1,..,cxj,..,xj,..,xn). By scaling and alternating, the last term is zero. **QED**

It is also useful to know how determinants behave on block matrices.
Lemma: If a matrix C has block form as below, then detC = detA.detB.
proof: Consider the block matrix C below, where both A and B are square matrices:

$$\begin{vmatrix} A & * \\ 0 & B \end{vmatrix} = C.$$ Then det(C) = detA.detB.

If B is not invertible then row reduction will introduce a row of zeroes at the bottom, so both B and the whole block matrix have det = 0. If B is invertible, then row operations of the third kind just discussed above, will change the matrix | * | to the zero matrix, since the rows of B are a basis for the space containing the rows of | * |. This does not change the determinant but changes C into a block matrix of the following form:

$$\begin{vmatrix} A & 0 \\ 0 & B \end{vmatrix},$$ which is a product matrix,

$$\begin{vmatrix} A & 0 \\ 0 & B \end{vmatrix} = \begin{vmatrix} A & 0 \\ 0 & I \end{vmatrix}\begin{vmatrix} I & 0 \\ 0 & B \end{vmatrix}.$$ Thus,

$$\det\begin{vmatrix} A & * \\ 0 & B \end{vmatrix} = \det\begin{vmatrix} A & 0 \\ 0 & B \end{vmatrix} = \det\begin{vmatrix} A & 0 \\ 0 & I \end{vmatrix} . \det\begin{vmatrix} I & 0 \\ 0 & B \end{vmatrix} = \det A.\det B,$$ by LaGrange's formula.

Cor: The determinant of a diagonal, or upper (or lower) diagonal matrix, equals the product of the diagonal entries.

Expansion along one column or row

The next result allows us to compute an nxn determinant as an alternating sum of (n-1)x(n-1) determinants. Suppose we want to expand along the first column. detA = det(A) = (Sum over s): sgn(s)∏j: a(s(j),j). Each term, i.e. each product ∏j: a(s(j),j) with fixed s, has one factor from the first column of A, namely a(s(1),1).
Thus these can be grouped into n subsets according to the value of s(1).
I.e. det(A) = (Sum, all s): sgn(s)∏j a(s(j),j) =

(Sum,s(1)=1), sgn(s)∏j: a(s(j),j)+...+ (Sum,s(1)=n), sgn(s)∏j: a(s(j),j).

Now every term in the first sum with s(1)=1, contains the factor a(1,1), which can be factored out, and similarly for the other sums. Thus we have: det(A) =

a(1,1)(Sum,s(1)=1): sgn(s)∏j≠1 a(s(j),j) +...+ a(n,1)(Sum,s(1)=n): sgn(s)∏j≠1 a(s(j),j).

Here the first partial sum, (Sum,s(1)=1): sgn(s)∏j≠1 a(s(j),j) is just the determinant of the (n-1) by (n-1) matrix obtained by omitting the first row and first column from A. But the second one, (Sum,s(1)=2): sgn(s)∏j≠1 a(s(j),j) is off by a sign, because in the matrix obtained by omitting the first column and second row of A, the n-1 columns are numbered 2,3,4,...,n, while the n-1 rows are numbered 1,3,...,n. Thus the permutation s with s(1) = 2, inducing the "identity" of these index sets takes 2to1, then 3to3, 4to4,..., n to n, and of course 1to2. This makes the sgn of this s = -1 instead of 1. So we have introduce a minus sign. In case s(1) = i, then s induces the "identity" map from the columns 2,3,...,n, to the rows 1,2,...,i-1,i+1,...,n if s(2)=1, s(3)=2,...,s(i) = i-1, and then s(i+1) = i+1,...,s(n)=n. So s differs from the actual identity permutation of {1,...,n} by i-1 transpositions. Thus each term (Sum,s(1)=i): sgn(s)∏j≠1 a(s(j),j), is equal to (-1)^(i+1) times the determinant obtained from A by omitting the 1st column and ith row.

This implies the determinant of A is a dot product of the first column of A with the vector of (±) the (n-1)x(n-1) determinants just described. Namely the element a(i,1) of the ith row in the first column is multiplied by (-1)^(i+1) times the (n-1)x(n-1) determinant of the matrix obtained by eliminating the first column of A and the ith row. If we denote by Aij the n-1 by n-1 matrix obtained from A by deleting the ith row and jth column, then we have detA = a(1,1)detA11 – a(2,1)detA21+ ......+(-1)^(n+1) detAn1.

In the same way, the determinant can be computed as a dot product with any column of A. If we use the jth column, the permutation with s(j)=i inducing the "identity" map from the columns (1,2,..,j-1,j+1,...,n) to the rows (1,2...,i-1,i+1,..,n) will differ from the identity on {1,...,n} by i-j transpositions, so we can multiply by (-1)^(i+j).

Thus for any fixed j: **detA =**
(-1)^(1+j)a(1,j)detA1j + (-1)^(2+j) a(2,j)detA2j + ...+ (-1)^(n+j) a(n,j)detAnj.
Since A and its transpose have the same determinant, we can compute detA = detA^t, by expanding along columns of A^t, i.e. along rows of A. Another approach to this formula is to check it yields a normalized alternating multilinear function, hence = detA.

It follows that if we form a matrix adj(A) whose rows are those (n-1)x(n-1) determinants, with appropriate signs, then the diagonal entries of the matrix product adj(A).A, will all equal det(A). The off diagonal entries moreover are the determinants of matrices having two equal columns, hence zero. We summarize this as follows.

Classical adjoint of a matrix (not the one from spectral theory)

Let Aij be the (n-1) by (n-1) matrix obtained by eliminating from A its ith row and jth column. The number (-1)^(i+j)det(Aij) is called the cofactor of the entry aij of A.

Denote by adj(A) the transpose of the matrix of cofactors of A, i.e. the (ij) entry of adj(A) is the cofactor (-1)^(j+i)det(Aji) of the element aji of A. Then we have:

Proposition (LaGrange-Cramer's rule): A.adj(A) = adj(A).A = ∂(A)Id.
proof: We defined adj(A) so that adj(A).A = ∂(A).Id. Then ∂(A).Id = ∂(A^t).Id = (∂(A^t).Id)^t = (adj(A^t).A^t)^t = A.(adj(A^t))^t = A.adj(A). **QED**

Cor: If ∂(A) ≠ 0, then (1/∂(A))adj(A).A = Id, in particular A is invertible.

Cor: (Cayley Hamilton) If ch(t) = det(t-A) is the characteristic polynomial of A, then ch(A) = [0], i.e. A satisfies its own characteristic polynomial.
proof: By Cramer's rule, ch(t).Id = det(t-A).Id = adj(t-A).(t-A). Thus in the ring of polynomials with matrix coefficients, the polynomial ch(t).Id is divisible by (t-A) from the right. Hence the value of ch(t) when A is substituted for t from the right is zero, by the non commutative factor theorem, with essentially the same proof as the commutative one. Indeed the coefficients of ch(t).Id are scalars from k hence commute with A, so in any sense, we have ch(A) = [0]. **QED.**

Remark: As an example of the non commutative remainder/factor theorem, let f(t) = Ct^n. Then right evaluation of f at A equals CA^n, hence f(t) -fright(A) = C(t^n – A^n) is right divisible by (t-A), since (t^n – A^n) = (t^n-1 + t^n-2 A+...+tA^n-2+A^n-1)(t-A). Hence f(t) = fright(A) + g(t)(t-A), so by uniqueness of divisibility by the monic (t-A), the remainder after right division by (t-A) equals the right value fright(A). Applying this argument to every term of a general f gives the result.

Remark: It took me about 40 years to understand this trivial proof of Cayley Hamilton from first principles of determinants. The only book I have seen it in, is Modern Higher Algebra, by A.A. Albert. It is detailed in my web notes for math 8000 and 845. The underlying reason it works is the isomorphism between the algebra of matrices with polynomial coefficients from k[t] and the algebra of polynomials in a commuting variable t, with coefficients from the non commutative ring of matrices over k. Thus the equation in Cramer's rule det(t-A).Id = (t-A).adj(t-A), holds both as matrices with polynomial entries and as polynomials with matrix coefficients.

Remark: The theory of determinants presented here goes through almost unchanged for matrices with entries in any commutative ring R. Only small modifications need be made, such as the statement that a matrix A is invertible iff detA ≠ 0, which becomes the statement that A is invertible iff detA is a unit in the ring R. I.e. almost all proofs depend only on addition and multiplication, few involve division.

References

A. Adrian Albert, Modern higher algebra.
Sheldon Axler, Linear algebra done right.
Paul Halmos, Finite dimensional vector spaces.
Insel, Friedberg, Spence, Linear algebra.
Chi Han Sah, Abstract algebra.
Shilov, Linear algebra.
Roy Smith, revised linear algebra, math 8000, 845, http://www.math.uga.edu/~roy/.
Sergei Treil, Linear algebra done wrong, http://www.math.brown.edu/~treil/.

These notes are a rewritten version of class lectures by Roy Smith