A primer of linear algebra

Chapter one: Linear spaces and linear maps

Linear algebra is about linear spaces, also called vector spaces, and linear maps between them. The first topic is therefore linear spaces.

Defn: A (real) **vector space** V is a set of "vectors" closed under addition, and under "scalar" multiplication of vectors by real numbers, and which is an "abelian group" under addition (the usual properties hold: associativity, commutativity and existence of a zero and negatives), and has the usual properties under scalar multiplication (multiplication by 1 acts as the identity, multiplication distributes over addition, a(bv) = (ab)v if a,b, are numbers and v is a vector).

The same definition can be given of a vector space with scalars in any field k, such as the complex numbers, or even the field E = k[X]/(f) where f is an irreducible polynomial over k.

Eg: The basic example is R^n = ordered n - tuples of real numbers, with component - wise operations: $(v_1,...,v_n) + (w_1,...,w_n) = (v_1+w_1, v_2+w_2,...,v_n+w_n)$ and $a(v_1,...,v_n) = (av_1,...,av_n)$. For any field k, k^n is a k vector space; as is Fun(S,k) = {k valued functions on a set S}.

Defn: A "**subspace**" of V is a non empty subset closed under addition and scalar multiplication.

E.g. C^∝(R) is the subspace of real valued functions on R which are infinitely differentiable.

Defn: Given a subspace W of V we define a new vector space **V/W**, the "quotient" of V by W, by identifying two vectors x,y in V provided x-y lies in W. Addition is defined by setting [v] + [w] = [v+w], where [v] denotes the equivalence class of v, and c[v] = [cv].

Ex. Addition is well defined in V/W, i.e. independent of choice of representatives.

E.g. Vectors are equivalent in R^2/{y=0} if they have the same y coordinate.

Defn: For any two vector spaces V,W we define a new space **VxW**, the "direct product" of V and W, consisting of all ordered pairs (x,y) with x in V and y in W. Addition and multiplication are done separately on components as in R^n = the product of n copies of the real numbers.

Defn: A map f:V-->W from V to W is **linear** if f(x+y) = f(x)+f(y) for all x,y, in V, and if also f(ax) = af(x) for all x in V and all real numbers a.

E.g. Df = f' is a linear map D: C^∝(R)--> C^∝(R), more generally (D-c)f = f'-cf is linear.

Defn: An **isomorphism** is a linear map with a linear inverse.

E.g. Rotation 90degrees c.c., (x,y)-->(-y,x) is an isomorphism R^2-->R ^2, (with inverse what?). sending (a1,...,an) in k^n, to f in Fun({1,...,n},k) with f(i) = ai, is an isomorphism.

Defn: If f:V-->W is a linear map, then **ker(f)** = {v in V: f(v) = 0}, and **Im(f)** = {w in W: w = f(v) for some v in V}. Note: ker(f)is a subspace of V, and Im(f) is a subspace of W.

E.g. ker(D) = {constant functions}, ker(D-c) = {a e^(ct), all a in R}. Im(D) = ?, Im(D-c)=?

Ex: **1)** A bijective linear map is an isomorphism.
2) Given V and a subspace W, the map q:V-->V/W sending v to [v], is linear with ker(q) = W.
3) Given V,W the map π:VxW-->V sending (x,y) to x is linear and ker(π) = {0}xW.
4) The set **Hom(V,W)** of all linear maps V-->W is closed under addition and scalar multiplication, where (f+g)(v) = f(v)+g(v) and (cf)(v) = c(f(v)), hence Hom(V,W) also forms a vector space.
5) Hom(R,V) is isomorphic to V, by the map sending f to f(1).
6) If Hom(V,R) = V^* is the "dual" vector space of V, sending f to the operation "preceding by f" defines a linear map Hom(V,W)-->Hom(W*,V*), i.e. f:V-->W goes to the map g-->(gof).
7) Sending each x in V to "evaluation at x" defines a linear map V-->V** = Hom(V*,R).

If f:V-->W is any linear map then:
8) f is constant on equivalence classes in V/ker(f).
9) f defines a linear map [f]: V/ker(f)--->W sending [v] to f(v).
10) [f] in 9) is always injective, and [f] is surjective if and only if f was, hence [f] is an isomorphism if and only if f was surjective.
11) A linear map [f] can be defined the same way on V/U, for any subspace U contained in ker(f), but [f] will not be injective unless U = ker(f).

E.g. [D]: C^∝(R)/{constants}--> C^∝(R) is an isomorphism, with what inverse?

Defn: A "linear combination" of the vectors $\{v_1,...,v_m,....\}$ is a vector w which is a finite sum of scalar multiples of the given ones, i.e. a vector of form $w = a_1v_1+...+a_mv_m$. The term also denotes the summation on the right.

Eg: In R^3, (4,-5,1) = (8,-1,7) - 2(2,2,3)) is a linear combination of (2,2,3) and (8,-1,7).
In the C-space of C-valued functions on R, e^(it) = cos(t) + i sin(t), is a lin. comb. of cos, sin.

Defn: A set S of vectors "**spans**" or "**generates**" a vector space V iff every non zero vector in V is a linear combination of vectors in S, or equivalently if the set S is not contained in any proper subspace of V. In particular, the empty set spans the space {0}.

Eg: The set {(1,0), (0,1)} spans R^2 since any vector (a,b) can be written as the linear combination a(1,0) + b(0,1) = (a,b). The set {cos(t), sin(t)} spans ker(D^2 + 1).

Ex: For any subset $S = \{v_1,...,v_m,....\}$ of a vector space V, the set of all finite linear combinations of the vectors in S, plus 0 (if S is empty), forms a subspace L(S) of V which is spanned by S.

Defn: A space V is **finite dimensional** if V has a finite spanning set.

Defn: An indexed set of vectors $\{v_1,...,v_m,....\}$ is **independent** if the only scalars $a_1,..,a_m$ such

that $a_1v_1+...+a_mv_m = 0$ are $a_1=...= a_m = 0$; i.e. if when any $a_i \neq 0$, then $a_1v_1+...+a_mv_m \neq 0$.

Eg. {(1,0), (0,1)} is independent since the only way we can have a(1,0) + b(0,1) = (a,b) = (0,0), is to have a = b = 0. The empty set is independent. {cos(t), sin(t)} is independent.

Ex: In a dependent set some vector is in the space spanned by the others. Vectors in a sequence are dependent iff some vector is in the space spanned by the previous vectors.

Defn: A subset S of V is a **basis**, if S is independent and L(S) = V.

Eg: The set of unit vectors $e_1 = (1,0,...,0),...., e_n = (0,....,0,1)$, is a basis of R^n called the "standard basis". The set {(3,0), (2,5)} is another basis of R^2.

Ex: If $\{v_1,...,v_n\}$ is a basis of V, and $\{w_1,...,w_m\}$ is a basis of W, then $\{(v_1,0),...,(v_n,0), (0,w_1),...,(0,w_m)\}$ is a basis of VxW.

Ex: i) A finite sequence $\{v_1,...,v_n\}$ in a k vector space V, defines a unique linear map $f:k^n$-->V sending $(a_1,...,a_n)$ to $a_1v_1+...+a_nv_n$.
ii) The map f in i) is injective if and only if $S = \{v_1,...,v_n\}$ is independent, and
iii) f is surjective if and only if $S = \{v_1,...,v_n\}$ spans V, and
iv) f is an isomorphism if and only if $S = \{v_1,...,v_n\}$ is a basis for V.

Cor: There is a one to one correspondence between linear maps $f:k^n$-->V and ordered subsets $\{v_1,...,v_n\}$ of n vectors in V, in fact Hom(k^n,V) and Fun({1,...,n},V) are isomorphic spaces. In particular V is finite dimensional/k iff for some n, there is a linear surjection k^n-->V.

Rmk: An (ordered) basis for V introduces linear coordinates into V, since by the isomorphism with R^n, a vector in V gets represented by a sequence of numbers, i.e. a coordinate vector in R^n.

Defn: An isomorphism V-->k^n is called a **coordinate system** for V, and an isomorphism k^n-->V is called a **parametrization** of V.

Def: If E is a ring containing the field k, E is a vector space over k, and if c is an element of E, k[c] = {space of polynomials in c}, denotes the span of the monomials {1, c, c^2, c^3,...}. c is called a "variable", or "transcendental/k" iff these monomials are independent/k. The symbol X is often reserved for a variable, and the infinite dimensional space k[X] = "the polynomial ring/k".

Eg: The subspace of polynomials of degree ≤ d, has basis the set of d+1 monomials $\{1,X,...,X^d\}$. In this basis the coordinates of $a_0 + a_1X +....+a_n X^n$, are its coefficients $(a_0, a_1,...., a_d)$. Another basis is the sequence $\{1, (1+X), (1+X+X^2),..., (1+X+...+X^d)\}$. Then the coordinate vector of 1 is (1,0,...,0), the coordinate vector of (1+X) is (0,1,0,...,0),..., and the coordinate vector of $(1+X+...+X^d)$ is (0,...,0,1).

Ex. The natural linear map k[X]-->Fun(k,k) with image = {space of polynomial functions on k}, is injective iff k is infinite. E.g. if k = Z/2, then X(X-1) goes to the zero function. Thus polynomials are not always the same as polynomial functions.

Thm: Every finite dimensional space V has a basis, i.e. V admits an isomorphism with some k^n.
Pf: Choose a finite spanning set S = $\{v_1,...,v_n\}$ of V. Throw out all zero vectors. If v_2 is a multiple of v_1, throw out v_2, if not keep it. If v_3 is a linear combination of $\{v_1,v_2\}$, throw out v_3, if not keep it. Continue throwing out vectors which are linear combinations of previous ones. Then the ones left are a basis. **QED.**

Cor: A basis S of V defines a one - one correspondence between linear maps from V to W and set functions from S to W, i.e. every function S-->W extends uniquely to a linear map V-->W.
Pf : This is true of k^n, hence of all finite dimensional spaces V. QED.

Cor: A linear surjection f:k^n-->V which is not injective, restricts to an isomorphism from some linear subspace k^m of k^n to V(where m < n).
Pf: The map f takes the standard basis of k^n to a generating set S for V. Reduce S to a basis B, and choose a subset T of standard basis vectors of k^n mapping bijectively to B, hence an isomorphism from the subspace L(T) of k^n, to V. L(T) is easily identified with k^m where m is the number of vectors in the subset T. **QED.**

Ex: If V = k^n and W is the subspace spanned by e_n, then V/W is isomorphic to k^{n-1}.

Defn: A space V has **dimension** = n, iff V is isomorphic to k^n, iff V has a basis of n vectors.

To show a space cannot have two different finite dimensions, we prove:
Thm: If k^n and k^m are isomorphic, then n = m.
Pf: (induction on n) There is no linear surjection f:k^1-->k^m if m >1, since the image vectors of f all have proportional entries. If $2 \leq n < m$ assume f is a linear surjection f: k^n --> k^m, and $\{e_1,...,e_n\}$, and $\{u_1,...,u_m\}$ are the standard bases of k^n and k^m. Then the composition k^n-->k^m/span(u_m) is surjective but not injective, since if $v \neq 0$ and f(v) = u_m, then v maps to [0] in k^m/span(u_m). Hence by previous Cor, k^n-->k^m/span(u_m) restricts to a surjection from some subspace k^k of k^n, with k < n, to k^m/span(u_m) = k^{m-1}. Since k < m-1, this contradicts the inductive hypothesis. **QED.**

Cor: Two finite dim'l spaces are isomorphic iff they have the same dimension.

Cor: All bases of a finite dimensional space have the same cardinality.

Convention: The space {0} has dimension zero; the empty set is a basis.

Thm: If W is a subspace of V and dim(V) < ∞, dimW + dim(V/W) = dimV.
Pf/Ex: Choose a basis $w_1,...,w_s$ for W, and extend it to a basis $\{w_1,...,w_s, v_1,...,v_t\}$ of V. Then $\{[v_1],...,[v_t]\}$ is a basis for (V/W). **QED.**

Thm: If f:V-->W is a linear surjection and dim(V) < ∞, then dim(ker(f))+dimW = dimV.
Pf: f induces an isomorphism from V/ker(f) to W. **QED.**

Cor: If V,W are finite dim'l, then dim(VxW) = dimV + dimW.
Pf: The projection taking (x,y) to y is a linear surjection from VxW to W with kernel Vx{0}, which is isomorphic to V. **QED.**

Lemma: If dim(V) < ∞, every independent set in V is contained in a basis.
Pf: If $\{v_1,...,v_n\}$ is independent, and $\{w_1,...,w_m\}$ is a basis, reducing the generating set $\{v_1,...,v_n,w_1,...,w_m\}$ to a basis, as above, does it. **QED.**

Ex: 1) If dim(V) = n, an independent set of vectors in V has ≤ n vectors.
2) If ∞ > dimV > dimW, no linear map V-->W is injective, and no linear map W-->V is surjective.
3) If $S = \{x_1,....,x_k\}$ are vectors in V, and dim(V) = n, then any two of the following implies the third. **a)** S is independent, **b)** S spans V, **c)** k = n.
4) If $T_1,....,T_k$:V-->V are surjective linear maps, with finite dimensional kernels, then the kernel of their composition $T_1o....oT_k$, has dimension equal to the sum $\dim.\ker T_1+....+ \dim.\ker T_k$.
5) A linear diff. operator with constant coefficients $L = (D-c_1)o....o(D-c_k)$: C^∝(R)--> C^∝(R), has k dimensional kernel. Give a basis for ker(L) and prove it is a basis.

Chapter Two: dot products, matrices, eigenvectors, and digonalizable linear maps.
Definition: The "dot product" of two vectors $(a_1,...,a_n)$ and $(b_1,...,b_n)$ in k^n is defined as: $(a_1,...,a_n).(b_1,...,b_n) = a_1b_1+....+a_nb_n$. It is a number. We say vectors a,b, in k^n are "orthogonal" iff a.b = 0.

The matrix of a linear map k^n --> k^m
Given a linear map $f:k^n \to k^m$, arrange the image vectors $f(e_1),...,f(e_n)$ as columns in a rectangular "matrix" A. Then there are m rows and n columns; we call A an "m by n" matrix. If $v = (a_1,...,a_n)$ is any vector in k^n, $f(v) = a_1f(e_1)+...+a_nf(e_n)$, is the linear combination of the columns of A having the coordinates of v as coefficients. Thus the ith entry of f(v) is obtained by dotting v with the ith row of A.
Thus f(v) can be computed by multiplying A by v as follows: write v as a length n column vector to the right of A. The product Av is a length m column vector, where the i^{th} entry of Av is the dot product of the ith row of A with v. Thus each linear map from k^n to k^m is represented by multiplying by a (unique) m by n matrix.

Eg: The matrix of the map $R^2 \to R^2$, f(v) = 6v, has rows (and columns): (6,0) and (0,6). The

matrix of the rotation map of R^2 counter clockwise through $\pi/2$ radians has columns (0,1), (-1,0).

Ex: Find the matrix of the reflection map of R^2 in the line spanned by (1,0), and the matrix for counter clockwise rotation about (0,0) through t radians.

Ex: i) The space of all m by n matrices forms a vector space Mat(m,n) where A+B is the matrix whose (i,j) entry, i.e. the entry in the ith row and jth column, is the sum of the (i,j) entries of A and B, and where cA is the matrix whose (i,j) entry is c times the (i,j) entry of A.
ii) The space Hom(k^n, k^m) is isomorphic to the space Mat(m,n), (note the indices n,m occur correctly in the reverse order here).
iii) The dimension of Mat(m,n), hence that of Hom(k^n,k^m), is mn.

The matrix associated to a linear map f:V-->W by bases of V,W.
If f is a linear map from one finite dimensional vector space V to another W, then by choosing bases for V and W we obtain isomorphisms between these abstract spaces and some coordinate spaces k^n and k^m, and hence a resulting linear map from k^n to k^m which has a matrix A. This A is called the matrix of f associated to the given bases for V and W. A map from V to itself has a matrix associated to any given basis of V.

If f is a linear map f:V-->W and $v_1,...,v_n$, and $w_1,...,w_m$ are bases of V,W, the jth column of the associated matrix for f, is composed of the coefficients $c_1,...,c_m$ where $f(v_j) = c_1w_1+...+c_mw_m$, is the unique basis expansion in W, for the image under f of the jth basis vector of V.

Eg: If D:V-->V takes a polynomial of degree ≤ 2 to its derivative, its matrix in the basis {1, X, X^2} has columns (0,0,0), (1,0,0), (0,2,0), since D(1) = 0 = 0(1,0,0) + 0(0,1,0) + 0(0,0,1), and D(X) = 1 = 1(1,0,0) +0(0,1,0) +0(0,0,1), and D(X^2) = 2X = 0(1,0,0) + 2(0,1,0) + 0(0,0,1).

Matrix multiplication corresponds to map composition.
Ex: If f:V-->W and g:W-->U are linear maps, and we choose bases for all three spaces, the matrix of the composition gof has as entry in its i^{th} row and j^{th} column, the dot product of the ith row of the matrix for g with the jth column of f. If A is the matrix of f, and B is the matrix for g, we write this matrix product as BA = the matrix for gof, (in the same bases).

Defn: An **eigenvector** of a linear map f, is a non zero vector v such that f(v) is a scalar multiple of v, i.e. such that f(v) = cv for some scalar c. The scalar c is the **eigenvalue** associated to v.

Geometry of eigenvectors: Recall that a vector v in R^3 say, has both a length and (if $v \neq 0$) a direction. An eigenvector is a non zero vector v such that either f(v) = 0, or v and f(v) have the same or opposite direction. Hence v spans a line that is mapped by f into itself.

Eigenvectors and diagonal matrices: If V has a basis consisting of eigenvectors for the map f:V-->V, i.e. an "eigenbasis", then the matrix A for f in this basis is "diagonal"; A has the eigenvalues of this basis on its main diagonal (upper left to lower right) and zeroes elsewhere.

Eg: The map c:V-->V multiplication by a scalar c, has diagonal matrix c.I in any basis. Thus every basis is an eigenbasis for the identity map.

Eg. The map f:R^2-->R^2 sending (1,0) to (1,0) and (0,1) to (0,2) has diagonal matrix with columns (1,0) and (0,2) in the standard basis, i.e. the standard basis is an eigenbasis. But (1,1) is not an eigenvector, so in the basis {(1,0), (1,1)}, the matrix of f has columns (1,0), (-1,2).

Ex: The derivative map D on real polynomials above, has no eigenbasis. More generally, if T:V-->V is nilpotent, i.e. if T^r = 0 for some r, but T (and V) is not zero, T has no eigenbasis.

The transpose of a matrix, symmetric matrices.

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

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

"Spectral theorem" (real symmetric matrices are orthogonally diagonalizable)

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

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

Alternate proof: Choose y in the unit sphere where f(x) = Ax.x has minimum value c, and set B = A-cI. Then for all x,t, $0 \leq B(y+tx).(y+tx) = 2t\ By.x+t^2\ Bx.x$. Hence By.x = 0 for every x in V, i.e. By = 0, so Ay = cy.

The minimal polynomial of a linear map (in finite dimensions)

Given a vector space V and a linear map f:V-->V, define a multiplication of the polynomial ring k[X] on V by saying that X times a vector v is f(v). similarly, X^2 times v is f(f(v)), and so on. Sending a polynomial P to multiplication by P, i.e. sending P to P(f), defines a linear map from k[X] to Hom(V,V). If dim(V) = n, then dim(Hom(V,V)) = n^2, but k[X] is infinite dimensional, with basis all monomials {1, X, X^2, X^3,....}. Thus the map k[X]-->Hom(V,V) has a non zero

kernel, i.e. for some $P \neq 0$, $P(f)=0$.

Defn: If f:V-->V is a linear map, the monic polynomial P of least degree such that $(P(f))v = 0$ for all v in V, is the **minimal polynomial** of f.

Ex: By the division algorithm, every polynomial in the kernel of the map k[X]-->Hom(V,V), is a multiple of the minimal polynomial of f.

Characterizing diagonalizabilty by the minimal polynomial.

Lemma: If T:V-->V has minimal polynomial $(X-c_1)^{r}1(X-c_2)^{r}2....(X-c_t)^{r}t$ with all c_i distinct, V is isomorphic to the product of $V_i = \ker(T-c_i)^{r}i$.

Pf: Consider the map from the product of the V_i to V, taking $(v_1,...,v_t)$ to $v_1+...+v_t$, which we must show is injective and surjective. Define polynomials $P_1,....,P_t$, where $P_1 = (X-c_2)^{r}2(X-c_3)^{r}3....(X-c_t)^{r}t$, $P_2 = (X-c_1)^{r}1(X-c_3)^{r}3....(X-c_t)^{r}t$, ..., $P_t = (X-c_1)^{r}1(X-c_2)^{r}2....(X-c_{t-1})^{r}t-1$. The Euclidean algorithm gives $Q_1,...,Q_t$ such that $P_1Q_1+...+P_tQ_t = 1$. Hence for any vector v in V, $v = P_1(T)Q_1(T)(v) +....+P_t(T)Q_t(T)(v)$ is in the sum of the images of the polynomials $P_i(T)$. Since $\mathrm{Im}(P_i(T))$ is in V_i, we have proved surjectivity.

For injectivity, assume $v_1+...+v_t = 0$. Then each v_i is a sum of the other v_j hence v_i lies in the kernel of P_i. But each v_i also lies in the kernel of every P_j with $j \neq i$, hence each v_i is in the kernel of $Q_1P_1+...+Q_t P_t = 1$, i.e. every $v_i = 0$. This proves injectivity. **QED.**

Thm: A linear map f:V-->V is diagonalizable iff its minimal polynomial is $(X-c_1)(X-c_2)....(X-c_t)$ where $t \leq \dim V$ and all c_i are distinct.

Pf: If f has a diagonal matrix in the basis $v_1,...,v_n$, with entries $c_1,...,c_n$ on the diagonal, note that the polynomial $(X-c_1)(X-c_2)....(X-c_n)$ does give zero when f is substituted for X, and this is still true when we omit repeated occurrences among the scalars c_i. On the other hand no proper factor of this reduced polynomial can annihilate all vectors in V, since if we omit say $X- c_1$, then the remaining factors map v_1 to $(c_1-c_2)....(c_1-c_t)\, v_1 \neq 0$. Thus $(X-c_1)(X-c_2)....(X-c_t)$, $t \leq n$, is the minimal polynomial.

Conversely if the map f has minimal polynomial as in the hypothesis, then V is isomorphic by the lemma to the product of subspaces $\ker(f-c_iI)$. Choosing bases of these subspaces and taking their union gives a basis for V consisting of eigenvectors for f. **QED.**

Eg: Since the map $f:R^2-->R^2$ with f(1,0) = (0,1), and f(0,1) = (0,0), satisfies $X^2 = 0$, i.e. f(f(v)) = 0 for all v, but f is not itself zero, f has minimal polynomial X^2. Hence by the theorem, f is not diagonalizable.

Ex: Prove directly that in the example f just above the only eigenvectors are (0,a). (Show f(v) = cv implies v = 0 or c = 0).

Nilpotent maps and Jordan canonical form

The lemma above showed that linear maps T whose minimal polynomials have linear factors with

multiplicities $(X-c)^r$, $r > 1$, lead to a decomposition involving subspaces on which the map T-c is not identically zero but is nilpotent, i.e. on which $(T-c)^r$ is identically zero. Consequently, on each factor space V_i in the lemma above, the operator T is the sum of the diagonal operator $c_i Id$ and the nilpotent operator $(T-c_i)$. Thus also on all of V, any operator T whose minimal polynomial has all its roots in the scalar field, is the sum of a diagonalizable operator D and a nilpotent operator N. Namely D is the unique operator on V whose restriction to V_i is $c_i Id$, and N is the unique operator whose restriction to V_i is $T-c_i Id$. This proves the following:

Cor: If all roots of the minimal polynomial of T lie in the scalar field, then T = D+N where D is diagonalizable, N is nilpotent, and DN = ND, and D,N are unique with these proeprties.

Ex: Prove that DN = ND, DT = TD, and NT = TN.

Pf of uniqueness in Cor: In the proof of the lemma above, note that $P_i(T)Q_i(T)$ is projection on V_i , i.e. is the identity on V_i and annihilates every other V_j. Since D is a linear combination of these projections, D is a polynomial in T, as is N = T-D. Thus if T = D' + N' is another decomposition where D', N' are respectively diagonalizable and nilpotent and commute, then D' and N' both commute with T as in the exercise. Thus D',N' commute with polynomials in T, hence with D and N above.

Ex: Conclude that N'-N is nilpotent.

Moreover D' commutes with projection on V_i, hence leaves V_i invariant. Since the minimal polynomial of the restriction of D' restricted to V_i divides that of D', it has distinct linear factors, so that restriction is diagonalizable. Then D'-D is diagonalizable on every V_i, hence on V. Thus if T = D+N = D'+N', then D'-D = N-N' is both diagonalizable and nilpotent, i.e. zero. Thus D = D', and N = N', i.e. D,N in the decomposition T = D+N, are unique. **QED.**

We will show next that an operator as above, whose minimal polynomial has all its roots in the scalar field, has a matrix almost as simple as a diagonal one, in a suitable basis.

Thm: With hypotheses as in the previous lemma, there exist bases in which the matrix of T is almost diagonal. I.e. each scalar c_i occurs $\dim V_i$ times on the diagonal, but there may also be 1's in some places just below the diagonal. [This is called a "Jordan" matrix for T.]
Pf: Since T commutes with any polynomial in T, each subspace V_i is invariant by T, i.e. $T(V_i)$ is contained in V_i, so it suffices to show each V_i has such a basis. Assume V is a space on which $(T-c)^r = 0$, but $(T-c)^{r-1} \neq 0$, i.e. $S = (T-c)$ is nilpotent on V of order r. Consider the quotient space $V/\ker(S^{r-1})$, and choose a basis $[x_1],...,[x_n]$ for it. S induces an injection from $V/\ker(S^{r-1})$ to the quotient $\ker(S^{r-1})/\ker(S^{r-2})$, hence we may extend the independent set $\{[S(x_1)],...,[S(x_n)]\}$ to a basis $\{[y_1],...,[y_{n+m}]\}$ for $\ker(S^{r-1})/\ker(S^{r-2})$. Continuing, we obtain a basis $\{[z_1],...,[z_{n+m+...+q}]\}$ for ker(S); then the set $\{x_1,...,x_n,y_1,...,y_{n+m},....,z_1,...,z_{n+m+...+q}\}$ is a basis for V.

(For independence, any relation among these, involving x's, would yield a relation among the $[x_i]$, a contradiction. Similarly if no x's occur, but some y's occur, we get a relation among the $[y_j]$, also a contradiction,... For spanning, count dimensions, using that dimV = the sum of the dimensions of the V_i.) Then this is the desired basis: $\{x_1,y_1,..,z_1; x_2,y_2,...,z_2 ; .... ; x_n,y_n,..,z_n ; y_{n+1},...,z_{n+1}; y_{n+2},...z_{n+2} ; y_{n+m},...,z_{n+m} ; ....; z_{n+m+....+1}, ... , z_{n+m+....+q}\}$.

Note each x_i is annihilated by S^r, each y_i is annihilated by S^{r-1}, ..., and each z_i is annihilated by S. Thus the z_i are the eigenvectors for T. Moreover S acts cyclically on these vectors in each block. I.e. $S(x_i) = y_i$, $S(y_i)$ = ... (we did not assign a letter to this one), ... , and so on down to $S(...) = z_i$. All the z_i belong to ker(S). Thus the r by r matrix for S acting on the block of basis vectors $\{x_i,y_i,...,z_i\}$ has first column (0,1,0,...,0), second column (0,0,1,0,....,0), 3rd column (0,0,0,1,0....,0), and r^{th} column all zeroes (0,....,0).

There are n blocks like this, then m blocks of size (r-1) by (r-1) corresponding to the blocks of basis vectors $\{y_{n+j},....,z_{n+j}\}$, and finally there are q blocks of size 1 by 1, i.e. one q by q block of all zeroes, corresponding to the remaining eigenvectors $\{z_{n+m+....+1}, ... , z_{n+m+....+q}\}$. Hence the matrix for T = S+cI, in this basis, is the same as just described, except also with c's everywhere on the diagonal. **QED.**

Eg: A linear map $T:R^2 \to R^2$ with minimal polynomial $(X-c)^2$ has Jordan matrix with columns (c,1), (0,c). The derivative D:V-->V on polynomials of degree at most 2, has minimal polynomial X^3, Jordan basis $\{X^2, 2X, 2\}$ and Jordan matrix with columns (0,1,0), (0,0,1), (0,0,0).

Cor: If the field of scalars is algebraically closed, e.g. complex numbers, the minimal polynomial has the form in the theorem, so every linear map has a Jordan matrix, (not always diagonal).

An infinite dimensional example: V = continuously differentiable functions on the real line, W = continuous functions. The derivative map D:V-->W, is linear and surjective by the fundamental theorem of calculus. The kernel of D is all constant functions by the mean value theorem. For any scalar c, $f(x) = e^{cx}$ is an eigenvector for D with eigenvalue c. The theory of Fourier series tries to approximate arbitrary functions by linear combinations of these eigenfunctions of D.

Ex: If $Lf = f^{(n)}+a_{n-1}f^{(n-1)}....+a_1f'+a_0f = 0$ is a linear differential operator with constant coefficients a_i, and if the characteristic polynomial $X^n+a_{n-1}X^{n-1}+....+a_1X+a_0$ of L factors into a product of distinct linear factors $(X-c_i)$, then the eigenfunctions $f(x) = e^{c_ix}$, for i = 1,...,n, are a basis of eigenvectors of the solution space $V = \{f: Lf = 0\}$. [For n = 1, and any solution f, $(f/e^{cx})' = 0$, so dimker(D-c) = 1. So $\ker((D-c_1)(D-c_2)...(D-c_n)) = (D-c_1)^{-1}(\ker(D-c_2)...(D-c_n))$ has dimension $\leq$ n. Then prove $\{e^{c_ix}: i = 1,..,n\}$, is independent.]

If the characteristic polynomial factors as a product of powers $(X-c_i)^{r_i}$, with some $r_i > 1$, there is no eigenbasis of $N(L) = \{f: Lf = 0\}$, but there is a Jordan basis $\{... ; e^{c_ix}, xe^{c_ix}, (1/2)x^2e^{c_ix},..., (1/(r_i-1)!)x^{r_i-1}e^{c_ix}; ...\}$. In particular, N(L) still has dimension n = degree of the polynomial.

Rational Canonical form, Cayley - Hamilton theorem.

The same ideas extend further when the minimal polynomial of a linear map T:V-->V factors as a product of (not necessarily linear) irreducible factors, say as $P_1^{r}1....P_n^{r}n$, where the P_i are distinct and irreducible in k[X], and k is the field of scalars. The same arguments show V is isomorphic to the product of the subspaces $V_i = \ker(P_i^{r}i)$, which are invariant under the action of T. If the power $r_i = 1$, we can view V_i as a vector space over the field $k[X]/(P_i)$, since the Euclidean algorithm for polynomials shows that one can divide in this quotient ring. I.e. k[X] is a k vector space, and the multiples of $P = P_i$ form a subspace, so the set of equivalence classes of polynomials k[X]/(P), (where two polynomials R,S are equivalent if P divides R-S), is a k vector space of dimension d = deg(P), with basis [1], [X], $[X^2]$,...,$[X^{d-1}]$. It is also a ring containing k as a subring, and since P is irreducible, for any polynomial R not divisible by P, we have RS +PQ = 1, for some polynomials S,Q. Hence in k[X]/(P) division by R is equivalent to multiplication by S, so k[X]/(P) is a field.

Recall dividing by $\neq 0$ scalars was the key to producing vector bases, so V is also a vector space over the field k[X]/(P) if the minimal polynomial P of T is irreducible. A basis for V over this field, consisting of s vectors, decomposes V into a product of s subspaces, each one d - dimensional over k and invariant under T, i.e. under multiplication by X. If $v \neq 0$ in V, the k[X]/(P) - subspace spanned by v, has k - basis {v, Tv, T(T(v)),..., $T^{d-1}(v)$}, corresponding to {1, X, X^2,....,X^{d-1}}. Since P(v) = 0, if $P(X) = a_0 + a_1X + a_2X^2+....+a_{d-1}X^{d-1} + X^d$, then $T^d(v) = -a_0v - a_1T(v) - a_2T^2(v)-....-a_{d-1}T^{d-1}(v)$, so the matrix of T in this k - basis has columns of form (0,1,0,...,0), (0,0,1,0,...,0), ..., $(-a_0, -a_1, -a_2, ..., -a_{d-1})$. Thus the matrix of T acting on a subspace with irreducible minimal polynomial P, and having dimension s over k[X]/(P), consists of s d-by-d blocks of that same form. Thus, if the minimal polynomial of T is a product of distinct irreducible factors P_i, the k - matrix of T in a suitable basis consists of a finite number s_i of such d_i-by-d_i blocks for each factor P_i.

Eg: If T:V-->V is a linear map on a real vector space, with minimal polynomial X^2+1, V is a sum of subspaces isomorphic to $R[X]/(X^2+1)$, i.e. of 2 dimensional real subspaces, each of them one - dimensional over the field $R[X]/(X^2+1)$. This quotient field is isomorphic to the complex numbers C, where X corresponds to i = sqrt(-1), so the operator T on this space corresponds to multiplication by i. If V has dimension s over C, the real rational canonical matrix of T on V, consists of exactly s blocks, and each block is 2 by 2, with columns (0,1), (-1,0).

Rmk: A rational canonical matrix always exists, composed of blocks like those above, but we must give a different proof for decomposability of the space when the minimal polynomial of T has repeated irreducible factors. We have already split our space as a product of subspaces $\ker(P_i^{r}i)$, but now the ring $k[X]/(P^r)$ operating on $\ker(P^r)$, is no longer a field, so we must still decompose each subspace $\ker(P^r)$ into "cyclic" T - invariant subspaces. E.g. if $V = \ker(P^r)$ where P is irreducible, T will always have a matrix of blocks as above, but the coefficients in the last column of each block will be those of some power P^s of P, with $s \leq r$.

The simplest case is $\dim(\ker(P^r)) = rd$ where $d = \deg(P)$. Then for any v with $P^r(v) = 0$ but $P^{r-1}(v) \neq 0$, a basis for V is given by $\{v, T(v),...,T^{dr-1}(v)\}$, and the k- matrix for T is the one block associated to the coefficients of P^r. Thus there is no problem when $\dim_k(\ker(P_i{}^{r_i})) = r_i.\deg(P_i)$ for all i, since no decomposition is necessary. In the general case, the spaces $\ker(P_i{}^{r_i})$ may be bigger. There is always a vector v in $\ker(P^r)$ with $P^{r-1}(v) \neq 0$, but then we must prove $\ker(P^r)$ splits as a product of the rd dimensional subspace spanned by $\{v, T(v),....,T^{dr-1}(v)\}$, and another T - invariant subspace. Then we can finish by induction on dimension. To get this splitting we can adapt the following result.

Splitting Lemma: If G is a finite abelian group of order p^r, p is prime, w an element generating a subgroup (w) of maximal order, and G/(w) = (z) is cyclic, then G is isomorphic to (w) x (z).
Pf: Assume f:G-->(z) sends y to z, and has kernel (w). If $\text{ord}(z) = p^a$, [$p^a z = 0$, but no smaller multiple = 0] then $\text{ord}(y) = p^{a+b}$, and $\text{ord}(w) = p^{a+b+c}$, where $a,b,c, \geq 0$. We seek $u = y+tw$ with $p^a u = 0$. Then f(u) = z, so $p^a \geq \text{ord}(u) \geq \text{ord}(z) = p^a$, so mapping z to u splits G as (w)x(u), isom. to (w)x(z). But $f(p^a y) = p^a z = 0$ so $p^a y$ is in (w), and no smaller multiple is, so $(y) \cap (w) = (p^a y)$ has order p^b, so $(p^a y) = (p^{a+c} w)$. So for some s, $p^a y = s p^{a+c} w = p^a (s p^c w)$, and $p^a(y - s p^c w) = 0$. So let $u = (y - s p^c w)$. **QED.**

Cor: If G is a finite abelian p group, ord(w) maximal, then G/(w) is a product of cyclic groups (z_i) by induction. If $f:G \to G/(w) = \prod(z_i)$ sends y_i to z_i, apply lemma to the map $f:(w,y_i) \to (z_i)$, with kernel (w). Mapping each z_i to the corresponding u_i in G splits G as (w) x $\prod(u_i)$.

Ex: Prove the splitting lemma needed for the general rational canonical form, with an irred. poly P replacing the prime integer p.

Eg: If T has minimal polynomial X^r, T has a matrix of blocks of size $\leq r$, with columns (0,1,0,...,0), (0,0,1,0,...,0), ..., (0,0,.....,0,1, (0,0,....,0). This is both the Jordan and rational canonical form of T.

Rmk: In the discussion of Jordan form, the Jordan matrix for T on the subspace V_i, equals c_iId plus the rational canonical matrix of the nilpotent operator $T - c_i$. So Jordan form is a special case of rational canonical form.

Summary review of determinants.
For actually calculating diagonal, Jordan, and rational canonical forms, we require the abiklity to compute minimal polynomials and eigenvalues. For this determinants are useful.

Definition: If A is an n x n matrix over R, define for each (i,j) with $1 \leq i,j \leq n$, let A_{ij} = the (n-1)x (n-1) matrix obtained by deleting from A the ith row and jth column. Then define determinants recursively as follows: If n=1, and A = (a) define D(A) = a. If we have defined D for all (n-1) x(n-1) matrices, and if A is in $\text{Mat}_n(R)$, set $D(A) = a_{11} D(A_{11}) - a_{12} D(A_{12}) \_ .... \_ a_{1n} D(A_{1n})$, (expansion by the first row).

Example: det |a b | = ad-bc. det | a b c | = a(ei-fh) – b(di-fg) + c(dh-eg).
|c d| |d e f |
|g h i |

Remark: We can expand determinants also by columns; i.e. for any choice of column, say j, we have $D(A) = \sum_i (-1)^{i+j} a_{ij} D(A_{ij})$.

Corollary: If A is upper or lower triangular, e.g. diagonal, then $D(A) = \prod a_{ii}$, the product of the diagonal entries. If A is a matrix with two equal rows or columns, then D(A) = 0.

Theorem: D(A) is n-linear and alternating as a function of the rows and columns of A. Hence,
i) If A' is the result of interchanging two rows or columns of A, D(A') = - D(A).
ii) If A' is the result of adding to one row (or column) of A, a scalar multiple of another row (or column), then D(A') = D(A).
iii) If A' is the result of multiplying a row or column of A by a scalar c, then D(A') = c.D(A).

Rmk: A matrix can be rendered upper triangular by repeating operations i) and ii).

Prop: (i) $D(A^*) = D(A)$, where A^* is the transpose of A.
(ii) If A,B are n x n matrices, then D(AB) = D(A)D(B); in particular if B is invertible then $D(B^{-1}AB) = D(A)$, so all matrices for the same linear map f:V-->V (i.e. wrt any basis of V), have the same determinant, so we can define the determinant of a linear map V-->V.
(iii) If A is a square matrix, then A is invertible if and only if $D(A) \neq 0$.
(iv) Given A, let $B = (b_{ij})$ be the matrix: $b_{ij} = (-1)^{i+j} D(A_{ji})$. (Note the interchange of indices.) Then AB = D(A)I = BA. Thus if $D(A) \neq 0$, then $D(A)^{-1}B$ is a (two sided) inverse for A.

Ex. Define T:k^3-->k^3 by T(e1) = (-1,1,1), T(e2) = (1,-1,1), T(e3) = (1,1,-1). Find the characteristic and minimal polynomials of T, a basis of eigenvectors for T, and a matrix B such that B^(-1)[T]B is diagonal, where [T] is the standard matrix for T.

The characteristic polynomial of a linear map on a finite dimensional space

Since all matrices for a linear map T:V-->V, in all bases, have the same determinant, define the "characteristic polynomial" of T as det(X.I-T). It follows from expanding about the last column that a rational canonical block matrix satisfies its own characteristic polynomial. In fact the map acts cyclically on all but the last vector, so no power of T smaller than the size of the block is a linear combination of smaller powers, so the polynomial associated to the last column of a rational canonical block is both the minimal and characteristic polynomial. Since every linear map T:V-->V has a rational canonical form, we obtain:

Thm(Cayley Hamilton): Every linear map T:V-->V on a finite dimensional space V, satisfies its characteristic polynomial, whose degree equals dim(V). In particular the minimal polynomial divides the characteristic polynomial, and hence has degree ≤ dim(V).

Examining the rational canonical decomposition, the characteristic polynomial of T equals

the product of the associated polynomials of every block in its rational canonical matrix, and the minimal polynomial equals only the product of the polynomials associated to the largest block for each irreducible factor. In particular the minimal and the characteristic polynomial have the same irreducible factors. So if one computes a determinant and can then factor the characteristic polynomial into irreducible factors P_i, one can explicitly compute canonical forms of a matrix by finding bases for the kernels of the powers $P_i^s(T)$, $s \leq r_i$.

Ex: A scalar c is an eigenvalue of T if and only if det(T-c.I) = 0, iff c is a root of the characteristic polynomial of T, if and only if c is a root of the minimal polynomial of T.

Ex: Find all Jordan and rational canonical forms of linear maps of R^3 with minimal polynomials $(X-2)^3$, $(X-2)^2$, and $(X-2)$.

Ex. Find the Jordan forms of these matrices:

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

Ex. Every linear map from R^3 to itself has at least one eigenvector.

Ex. A linear map of R^3 that preserves dot products is composed of rotations and reflections.

Ex. Give a detailed proof of Cayley - Hamilton.