845 notes, part one; "Linear Algebra": Modules and Matrices

Introduction: We have seen that one can study field extensions very profitably by means of their Galois groups of automorphisms, and in particular that the Galois group of an extension of k of degree n is isomorphic to a subgroup of the "linear" group $Aut_k(k^n)$ of isomorphisms of the vector space k^n. In this part of the course we study self homomorphisms, or "endomorphisms" of arbitrary vector spaces over fields and of modules over rings. These endomorphisms are no longer always isomorphisms, hence form a ring rather than a group, but it is again fundamental to classify them up to "similarity" (the analog of conjugacy). The extension from a field to a more general ring acting on an abelian group adds greatly to the applications of the method. For instance an abelian group which admits an action by the polynomial ring k[X] is equivalent to a pair (V,T) consisting of a k vector space V and a linear transformation $T:V \to V$. Thus the generalization lets us analyze both T and V.
In particular, a module theoretic generalization of the decomposition theorem for finite abelian groups will allow us to classify linear transformations of finite dimensional vector spaces, especially those over $\mathbb{C}$, in terms of their "rational", or "jordan" normal forms. These canonical forms provide good representatives for conjugacy classes in the linear group $Aut_k(k^n)$, and of similarity classes in the ring $End_k(k^n)$, hence computing them makes it possible to decide when two elements of $End_{\mathbb{C}}(\mathbb{C}^n)$ are similar to each other. In particular we give several criteria, called "spectral theorems", for a matrix to be similar to the simplest possible matrix, a diagonal matrix.
Vector space and module theoretic methods are fundamental tools in differential equations, geometry, and topology. Within the field of algebra, "linear groups" over finite fields provide new examples of finite simple groups.

Acknowledgement: The late Maurice Auslander, who taught me about functors and modules, had a way of making the most abstract constructions seem natural which I hope to emulate in these notes.

Terminology: In these notes "non abelian" or "non commutative" may mean "not necessarily abelian", or not necessarily commutative".

Suggestion: This is a good time to go back and reread sections §10, and §15 in the 843 notes, on free abelian groups and vector spaces.

§1) Fundamental theorem of finite abelian groups revisited, via matrices and linear maps.

Decomposing finite abelian groups into a product of cyclic groups last quarter seemed as if it should have been easy, but it required a slightly tricky argument to eliminate overlaps between the cyclic subgroups which generate a given group. This process of transforming given generators into "independent" ones is unavoidable, so we want to give a systematic procedure to simplify it. The new method will exploit the Euclidean algorithm, and will simultaneously prove the decomposition theorem for finitely generated, not necessarily finite, abelian groups, and for finitely generated "modules" over any Euclidean domain. A slight variation yields the decomposition theorem, but not a computational procedure, also over any p.i.d. These generalizations will give easy proofs of the standard existence theorems for normal forms of matrices for linear transformations of vector spaces, (but will not render easy their actual calculation).

We will begin to write most of our abelian groups additively, to make more natural the analogy between abelian groups and vector spaces. Recall that if $(M,+)$ is an abelian group, then it is possible to multiply elements of M by integers, where for x in M, and n in $\mathbf{Z}^+$, $nx = x + x + \ldots + x$ (sum of n terms), and $(-n)x = n(-x) = -x - x \ldots - x$ (n terms). Thus there is a natural action on M, by the ring $\mathbf{Z}$. This multiplication is distributive over addition in M, hence gives a group homomorphism of M to itself, and the map taking an integer n to "multiplication by n" defines a map $\mathbf{Z} \to \mathrm{End}(M)$. Since multiplication is also distributive over addition in $\mathbf{Z}$, $[(n+m)x = nx+mx]$, and associative, $[(nm)x = n(mx)]$, this map $\mathbf{Z} \to \mathrm{End}(M)$, which takes n to $n \cdot \mathrm{id}$, is a ring map. So viewing an abelian group M as a $\mathbf{Z}$-module means exploiting the action on M of the subring of End(M) generated by the identity endomorphism. Advantages of this approach include the use of integer matrices to represent homomorphisms, and the realization that other more exotic rings, such as the non commutative ring End(M), may act interestingly on M as well.

If $x_1, \ldots, x_n$ are elements of M, a "linear combination" (or a $\mathbf{Z}$-linear combination) of these elements is a finite expression $\Sigma \alpha_i x_i$, with α_i

in $\mathbf{Z}$. This is an element of M. When we use just the word "map" for a function between two groups you may assume it is a group homomorphism; similarly a "map" between two rings is a ring homomorphism. We will sometimes denote the product $\Pi_{j=1,...,m}$ A of m copies of the ring A, by A^m.

Definition: An abelian group M is "finitely generated" iff there is a finte subset $S = \{x_1,...,x_n\}$ of M such that every element of M is a linear combination (not necessarily unique) of the elements of S.

Definition: A cyclic group is a group which is isomorphic to a quotient of $\mathbf{Z}$ by a subgroup.
Note: This is equivalent to our old definition.

Definition: A subset S of $\mathbf{Z}^n$ is called a "basis" or "free basis" iff each element y of $\mathbf{Z}^n$ has a unique expression as $\Sigma\alpha_j x_j$ with α_j in $\mathbf{Z}$.

Recall: An abelian group M is free on $S \subset M$ iff every set function $S \to G$, from S to an abelian group G, extends uniquely to a group homomorphism $M \to G$.

Exercise #121) Prove: $\mathbf{Z}^n$ is "free" on a subset $S = \{x_1,...,x_n\} \subset \mathbf{Z}^n$ iff S is a basis for $\mathbf{Z}^n$.

Exercise #122) Prove: A subset $\{x_1,...,x_n\} \subset \mathbf{Z}^n$ is a basis iff there is an isomorphism $\varphi: \mathbf{Z}^n \to \mathbf{Z}^n$ with $\varphi(e_j) = x_j$, for all j.

Fundamental theorem of finitely generated abelian groups:
Theorem: A non trivial finitely generated abelian group M is isomorphic to a unique finite product of cyclic groups $M \cong (\mathbf{Z}^r \times \Pi_j (\mathbf{Z}/n_j))$ where for all i (if any exist), $n_i \geq 2$ and $n_i | n_{j+1}$.

Remark: The uniqueness means that the sequence of integers n_i and the number r of factors of $\mathbf{Z}$ are uniquely determined by M. The number r is called the "rank" of M, and the integers n_j are called the "invariant factors" of M. Thus two finitely generated abelian groups are isomorphic iff they have the same rank and the same invariant factors. The rank of M should be thought of as

analogous to the dimension of a (finite dimensional) vector space, but the invariant factors are something a vector space does not have, because unlike $\mathbf{Z}$, a field has no interesting quotients.

First we prove the uniqueness, assuming a decomposition exists.
Lemma: Two finite products of infinite cyclic groups are isomorphic, $\mathbf{Z}^m \cong \mathbf{Z}^r$, if and only if $m = r$.
proof: If $M = \mathbf{Z}^m$, then $2M = (2\mathbf{Z})^m$, and thus $M/2M \cong (\mathbf{Z}/2)^m$ has order 2^m where m is the number of copies of $\mathbf{Z}$ in the product M. Hence if $M = \mathbf{Z}^m \cong \mathbf{Z}^r = N$, then $M/2M \cong N/2N$, so these two finite groups have the same order, $2^m = 2^r$. Thus $m = r$. QED.

Corollary: Rank(M) is well defined at least if $M \cong \mathbf{Z}^r$.

Definition: If M is an abelian group, let $t(M) \subset M$ be the "torsion subgroup", $t(M) = \{$those x in M such that $nx = 0$ for some $n \neq 0\}$. Thus x is in $t(M)$ iff x generates a finite cyclic subgroup $(x) \subset M$.

Note: If $M = (\mathbf{Z}^r \times \Pi_i (\mathbf{Z}/n_i))$, then $t(M) \cong \Pi_i (\mathbf{Z}/n_i)$, and $M/t(M) \cong \mathbf{Z}^r$. Thus rank(M) = rank(M/t(M)) = rank($\mathbf{Z}^r$) = r. This proves the rank of M is well defined for any finitely generated M. Since M determines $t(M) \cong \Pi_i (\mathbf{Z}/n_i)$, the uniqueness of the invariant factors of M follows from the earlier theorem on the uniqueness of the decomposition of the finite abelian group t(M). By the earlier theorem, instead of requiring n_i to divide n_{i+1}, we can choose the moduli n_i to equal $p_i{}^{r_i}$, for some (not necessarily distinct) primes p_i, and again we have uniqueness of the n_i.
QED. for uniqueness.

Remarks on what is unique and what is not: We have said that in the decomposition $M \cong (\mathbf{Z}^r \times \Pi_{i=1\ldots s} (\mathbf{Z}/n_i))$, it is the integers r and (n_i) which are unique, provided $n_i | n_{i+1}$. The choice of isomorphism itself however is not unique, in particular there are no distinguished subgroups of M corresponding to the individual factors of the product. In fact there is no special subgroup of M isomorphic to $\mathbf{Z}^r$, i.e. there is no distinguished subgroup of M corresponding to the "free part" of M. For example, the map $\varphi: \mathbf{Z} \times (\mathbf{Z}/2) \to \mathbf{Z} \times (\mathbf{Z}/2)$ where $\varphi(a,[b]) = (a,[a+b])$ is an isomorphism with inverse $\varphi^{-1}(a,[b]) =$

(a,[b-a]). Thus the subgroup $\mathbf{Z}\times\{0\}$ could be replaced by the subgroup $\mathbf{Z}\cdot(1,[1])$ in the decomposition of $M = \mathbf{Z}\times(\mathbf{Z}/2)$. Hence either $\mathbf{Z}\times\{0\}$ or $\mathbf{Z}\cdot(1,[1])$ can be thought of as the "free part". Another way to see this is that the map $\pi:\mathbf{Z}\times(\mathbf{Z}/2)\to\mathbf{Z}$ where $\pi(a,[b]) = a$, has more than one right inverse. I.e. in addition to the obvious choice $a\mapsto(a,[0])$, $\varphi(a) = (a,[a])$ is also a right inverse of π. From our general theory of splitting maps, every right inverse gives a splitting, so there is more than one way to split off a copy of $\mathbf{Z}$ inside of $\mathbf{Z}\times(\mathbf{Z}/2)$. Thus in the decomposition $M \cong (\mathbf{Z}^r \times \prod_{i=1,\dots,s} (\mathbf{Z}/n_i))$, neither the individual copies of $\mathbf{Z}$, nor the whole "free factor" $\mathbf{Z}^r$ correspond to distinguished subgroups of M.

What about the finite factors? Recall that in a finite abelian group, any element of maximal order splits off a cyclic summand, so in $(\mathbf{Z}/2)\times(\mathbf{Z}/4)$ we could split off the subgroup generated by ([1],[1]) as the copy of $\mathbf{Z}/4$, instead of using the subgroup $\{[0]\}\times(\mathbf{Z}/4)$ generated by ([0],[1]). Thus in our decomposition $M \cong (\mathbf{Z}^r \times \prod_{i=1,\dots,s} (\mathbf{Z}/n_i))$ there is no special subgroup of M isomorphic to $\mathbf{Z}/n_s$, for instance, assuming $s > 1$. However the entire "torsion part" $\prod_{i=1,\dots,s} (\mathbf{Z}/n_i)$ of the product does correspond to a unique subgroup of M, namely the "torsion subgroup" t(M) = {elements of finite order in M}. Thus there is at least one distinguished subgroup of M independent of choice of decomposition, namely t(M) = the largest finite subgroup of M.

There are in general however, other special subgroups of M which are naturally determined, independent of choice of decomposition, namely the Sylow subgroups M_p of t(M). The subgroup t(M) breaks down naturally further into the product of the subgroups M_p, although these do not occur in the standard decomposition above of $t(M) \cong \prod_{i=1,\dots,s} (\mathbf{Z}/n_i)$ where $n_i|n_{i+1}$. As remarked last quarter, one has two choices in decomposing a finite abelian group: one can either make the standard decomposition of the group as above, or one can first make a preliminary decomposition of the group into a product of Sylow subgroups, and then make the standard decomposition of each Sylow subgroup separately. The second approach is sometimes called the "prime power" decomposition.

To summarize, a finitely generated abelian group M has a natural, uniquely determined, torsion subgroup t(M) such that M/t(M) is a

free abelian group. M has many free abelian subgroups N isomorphic to the quotient M/t(M) and such that $N \times t(M) \cong M$, but there is no natural choice of $N \subset M$. The torsion subgroup t(M) has a unique sequence of "invariant factors" $(n_1,....n_s)$ such that $t(M) \cong \Pi_{i=1....s} (\mathbf{Z}/n_i)$, but there are many choices of the actual subgroups $M \supset N_i \cong \mathbf{Z}/n_i$ of M corresponding to the factors in this decomposition. The torsion subgroup has a unique collection of Sylow subgroups $M_p \subset t(M)$, for primes p dividing $\#(t(M))$ such that $t(M) \cong \Pi_p M_p$. These Sylow subgroups are not generally cyclic however, so to decompose t(M) further into a product of cyclic subgroups, one must make a standard decomposition of each M_p. For each M_p, there are in general many ways again to do this. I.e. although the orders of the cyclic subgroups of M_p which correspond to the factors in a standard decomposition are always the same, the subgroups themselves can be chosen in many ways.

Proof of existence of the decomposition:
Since M is finitely generated, exercise #112 below implies there is a surjective map $\sigma: \mathbf{Z}^m \to M$ which has a kernel $K = \ker(\sigma) \subset \mathbf{Z}^m$, such that $M \cong \mathbf{Z}^m/K$. We need to know that K is also finitely generated, but it is easy to prove something stronger.

Lemma: If $K \subset \mathbf{Z}^m = M$ is a subgroup of a finitely generated free abelian group, then K is also free, finitely generated, and $\text{rank}(K) \leq \text{rank}(M)$.
proof: We use induction on the number of factors. To get started, it is crucial that $\mathbf{Z}$ is a p.i.d. I.e. if $K \subset \mathbf{Z}$, then either $K = (0)$ or $K = n\mathbf{Z} \cong \mathbf{Z}$, for some $n \neq 0$. So $K \subset \mathbf{Z}$ is free of rank 0 or 1. Now assume the theorem for a product of at most m-1 factors, let $M = \mathbf{Z}^m$, and $K \subset M$. We try to split K between the first m-1 factors and the last one. I.e. consider the projection map $\mathbf{Z}^m \to \mathbf{Z}$, taking $(x_1,....,x_m) = x_m$, and denote by σ its restriction to K, $\sigma: K \to \mathbf{Z}$. Then $\text{Im}(\sigma) = \sigma(M) \subset \mathbf{Z}$ is either (0) or $\cong \mathbf{Z}$, and we have a surjective map $\sigma: K \to \text{Im}(\sigma)$. If $\sigma(K) = (0)$ we are done by induction since then K is a subgroup of the product of the first m-1 factors, so we may assume $\sigma(K) \cong \mathbf{Z}$, and that we have a surjection $\sigma: K \to \mathbf{Z}$.

Claim: The map σ splits, so that $K \cong \ker(\sigma) \times \mathbf{Z}$.
proof: By our general splitting criterion (lemma 3, §18, 844 part 2),

we only need a right inverse to σ, i.e. a group map $\varphi: \mathbf{Z} \to K$ such that $\sigma \circ \varphi = id_{\mathbf{Z}}$. Such a map φ is extremely easy to define because $\mathbf{Z}$ is such a nice group, i.e. a free abelian group. Since σ is surjective, we just pick any element x of K with $\sigma(x) = 1$, and define $\varphi(1) = x$. This extends uniquely to a group map $\varphi: \mathbf{Z} \to K$ with $\varphi(n) = nx$. Since $(\sigma \circ \varphi)(1) = \sigma(\varphi(1)) = \sigma(x) = 1$, we have $(\sigma \circ \varphi)(n) = \sigma(\varphi(n)) = \sigma(nx) = n \cdot \sigma(x) = n$, for all n. QED claim.
Now since σ splits, we have $K \cong \ker(\sigma) \times \mathbf{Z}$, and since $\ker(\sigma)$ is a subgroup of $\mathbf{Z}^{m-1}$, by induction $\ker(\sigma)$ is $\cong$ to a product of $\leq m-1$ copies of $\mathbf{Z}$. Thus $K \cong \ker(\sigma) \times \mathbf{Z}$ is $\cong$ to a product of $\leq m$ copies of $\mathbf{Z}$. **QED lemma.**

Exercise #123) If M is an abelian group with a finite set of generators $(x_1,\ldots,x_m)$, prove:
(i) there is a surjective group map $\sigma: \mathbf{Z}^m \to M$;
(ii) the image N of a surjective map $\varphi: M \to N$ is also finitely generated with at most m generators;
(iii) any subgroup $N \subset M$ is also finitely generated with at most m generators.

Now back to the decomposition theorem: we have a surjective map $\sigma: \mathbf{Z}^m \to M$, with kernel K, so that $M \cong \mathbf{Z}^m/K$, and $K \cong \mathbf{Z}^n$, with $n \leq m$. We want to conclude that M is a product of cyclic groups. There is one case where this is easy, namely when the isomorphism $\mathbf{Z}^n \to K \subset \mathbf{Z}^m$ takes each factor of $\mathbf{Z}^n$ into a different one of the factors of $\mathbf{Z}^m$, i.e. when K is a "subproduct" of $\mathbf{Z}^m$, as in the following:

Lemma: If $B = \Pi_j \mathbf{Z}$ is free abelian of finite rank m, $(r_1,\ldots,r_m)$ is a sequence of m non-negative integers, and $B \supset K = \Pi_j (r_j\mathbf{Z})$, then $B/K \cong \Pi_j (\mathbf{Z}/r_j)$, (where $\mathbf{Z}/0 \cong \mathbf{Z}$).
proof: The canonical map $B = (\Pi_j \mathbf{Z}) \to \Pi_j (\mathbf{Z}/r_j)$ taking $(x_1,\ldots,x_m)$ to $([x_1],\ldots,[x_m])$ has kernel $K = \Pi_j (r_j\mathbf{Z})$. Hence the result follows from the first isomorphism theorem. **QED.**

Now if all we know is $\sigma: \mathbf{Z}^m \to M$ is surjective, and $\ker(\sigma) \cong \mathbf{Z}^n$, how far away are we from the situation in the previous lemma? To see, let $\varphi: \mathbf{Z}^n \to \ker(\sigma)$ be the isomorphism in the previous sentence, and let $\varphi: \mathbf{Z}^n \to \mathbf{Z}^m$ be the composition $\varphi: \Pi_j \mathbf{Z} \to \ker(\sigma) \subset \Pi_j \mathbf{Z}$. Then we

have $M \cong \mathbf{Z}^m / \ker(\sigma) = \mathbf{Z}^m / \mathrm{Im}(\varphi)$, so at least M is completely determined by the map φ, i.e. by its target and image. Thus to understand M it suffices to understand φ well.
Claim: (i) φ can be represented by a matrix of integers, and (ii) the situation in the lemma occurs when that matrix is diagonal.

The matrix of a homomorphism of free abelian groups.
Let $\varphi: A = \mathbf{Z}^n \to B = \mathbf{Z}^m$ be a homomorphism between two free finite rank abelian groups, and let $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0)$, ..., $e_n = (0,\ldots,0,1)$, be the "standard basis" of the source group A. Note that each element x of A has a unique representation as a linear combination $\Sigma \alpha_i e_i$ with α_i integers. Since φ is a homomorphism, if $x = (\alpha_1,\ldots,\alpha_n) = \Sigma \alpha_i e_i$, then $\varphi(x) = \Sigma \alpha_i \varphi(e_i)$, so $\varphi(x)$ is determined if we know where φ sends the basis elements e_i. Since each element $\varphi(e_i)$ has a unique expansion as an "m-vector" in B, φ is thus determined by the sequence of mn integers representing the vectors $\varphi(e_1), \ldots, \varphi(e_n)$. It is traditional and practical to arrange these integers in an $m \times n$ (read "m by n") matrix (i.e. m rows and n columns), whose first column is the m-vector $\varphi(e_1)$, second column is the m-vector $\varphi(e_2)$, etc. We refer to the entry located in the ith row and jth column of a matrix as the (i,j) entry.

Remember: The jth column of the matrix $[\varphi]$ for φ is the image vector $[\varphi(e_j)]$ of the jth basis vector e_j of the source group A.

$$[\varphi] = [\,[\varphi(e_1)]\,[\varphi(e_2)]\ldots\ldots[\varphi(e_n)]\,]$$

Example: Suppose $\varphi: \mathbf{Z} \times \mathbf{Z} \to \mathbf{Z} \times \mathbf{Z} \times \mathbf{Z}$, is defined by $\varphi(1,0) = (1,2,3)$, and $\varphi(0,1) = (-2, 4, 0)$. Then the matrix is the following 3×2 array:

$$[\varphi] = \begin{bmatrix} 1 & -2 \\ 2 & 4 \\ 3 & 0 \end{bmatrix}$$

To compute $\varphi(a,b)$ using this matrix, write (a,b) as a column vector on the right of the matrix, and form the three "dot products" of the three rows with the vector (a,b). I.e. if (c,d) is a row of the matrix $[\varphi]$, recall the dot product with (a,b) is $(c,d)\cdot(a,b) = ca+db$. Thus we

get the following matrix product for $\varphi(a,b)$:

$$\varphi(a,b) = \begin{bmatrix} 1 & -2 \\ 2 & 4 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a-2b \\ 2a+4b \\ 3a \end{bmatrix} = (a-2b,\ 2a+4b,\ 3a).$$

Notice that the subgroup $\mathrm{Im}(\varphi)$ = {all elements of form $\varphi(x)$} = {all elements of form $\Sigma \alpha_j \varphi(e_j)$} = {all linear combinations of the columns of $[\varphi]$}. So in our case $\mathrm{Im}(\varphi)$ = the subgroup of $\mathbf{Z}\times\mathbf{Z}\times\mathbf{Z}$ spanned by (1,2,3) and (-2,4,0). Now the point of all this is to notice how the matrix looks in the situation of the previous lemma, where $B = \Pi_j \mathbf{Z}$ is free abelian (say of rank 3), (r, s, t) is a sequence of 3 non-negative integers, and $B \supset K = r\mathbf{Z} \times s\mathbf{Z} \times t\mathbf{Z}$. Then the natural isomorphism $\mathbf{Z}\times\mathbf{Z}\times\mathbf{Z} \to K$ is the map $(a,b,c)\mapsto(ra,sb,tc)$, taking e_1 to (r,0,0), e_2 to (0,s,0), and e_3 to (0,0,t), and thus with matrix:

$$[\varphi] = \begin{bmatrix} r & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & t \end{bmatrix}$$

So in this case the matrix is diagonal. The thing to note is that whenever the matrix is diagonal, then we are in the good case, as in the previous lemma. I.e. the image of φ, which is the subgroup $K \subset B = \Pi_j \mathbf{Z}$, is always the span of the columns, and when $[\varphi]$ is diagonal with diagonal entries r_j, the jth column is just $r_j e_j$, so then the subgroup $K = \Pi_j\, r_j\mathbf{Z}$ is a subproduct of $B = \Pi_j \mathbf{Z}$. In this case the lemma guarantees that the quotient $B/K \cong \Pi_j\, (\mathbf{Z}/r_j)$, is a product of cyclic groups. In general of course the matrix is not going to be diagonal, but we want to show that we can "change bases" in both the source and the target, from the standard bases (e_j) to possibly some other bases, thus composing φ with isomorphisms at both ends, in such a way that the composed map $\tilde{\varphi}: \Pi_i \mathbf{Z} \cong \Pi_i \mathbf{Z} \to \Pi_j \mathbf{Z} \cong \Pi_j \mathbf{Z}$, has a diagonal matrix. Then we can conclude that $\Pi_j \mathbf{Z}/\mathrm{Im}(\varphi) \cong \Pi_j \mathbf{Z}/\mathrm{Im}(\tilde{\varphi}) \cong$ a product of cyclic groups. The procedure we use will also show that the diagonal entries can be made to divide one another, and thus give the product decomposition in the standard form. [It is not necessary for the two free abelian groups, the source and the target, to have the same number of factors. If not square, a "diagonal" matrix will mean a diagonal square matrix plus some

extra zero rows at the bottom, or some extra zero columns on the right. In either case the columns still generate a subproduct of the target, which is the good case of the previous lemma.}

Now let's prove the key assertion made in this discussion.
Lemma: Let $\varphi: A \to B$ be a homomorphism of abelian groups, let $\sigma: \tilde{A} \to A$, and $\tau: B \to \tilde{B}$ be isomorphisms, and let $\tau \circ \varphi \circ \sigma = \tilde{\varphi}: \tilde{A} \to \tilde{B}$ be the composition. Then $B/Im(\varphi) \cong \tilde{B}/Im(\tilde{\varphi})$.
Proof: Composing $\tau: B \to \tilde{B}$ with the canonical map $\tilde{B} \to \tilde{B}/Im(\tilde{\varphi})$ gives a map $f: B \to \tilde{B}/Im(\tilde{\varphi})$, which is surjective since it is the composition of two surjections. Ker(f) is the subgroup of B that maps via τ into $Im(\tilde{\varphi})$. I.e. x is in Ker(f) iff $\tau(x) = \tilde{\varphi}(z) = (\tau \circ \varphi \circ \sigma)(z)$ for some z. We claim $Ker(f) = Im(\varphi)$. To see it, if x is in Ker(f), then $\tau(x) = \tilde{\varphi}(z) = (\tau \circ \varphi \circ \sigma)(z)$ for some z, so if we set $y = \sigma(z)$, then $\tau(x) = \tau \circ \varphi(y)$, and since τ is an isomorphism, $x = \tau^{-1} \circ \tau \circ \varphi(y) = \varphi(y)$. Thus x is also in $Im(\varphi)$, so $Ker(f) \subset Im(\varphi)$. On the other hand, if x is in $Im(\varphi)$, then $x = \varphi(y)$ for some y in A. Since σ is an isomorphism, $y = \sigma(z)$ for some z, and so $x = \varphi \circ \sigma(z)$, whence $\tau(x) = (\tau \circ \varphi \circ \sigma)(z) = \tilde{\varphi}(z)$, so that x does belong to Ker(f), and $Im(\varphi) \subset Ker(f)$. Thus $Im(\varphi) = Ker(f)$, hence by the first isomorphism theorem $f: B \to \tilde{B}/Im(\tilde{\varphi})$ induces an isomorphism $B/Im(\varphi) \cong \tilde{B}/Im(\tilde{\varphi})$. QED.

Exercise #124) In the setup of the previous lemma, prove that $\ker(\varphi) \cong \ker(\tilde{\varphi})$, and $Im(\varphi) \cong Im(\tilde{\varphi})$.

In our application of the lemma we can take $\tilde{A} = A$, and $\tilde{B} = B$.
Corollary: If $A = \mathbf{Z}^n$, $B = \mathbf{Z}^m$ (finite products), and $\varphi: A \to B$ is a homomorphism such that $M \cong B/Im(\varphi)$, and if there exist isomorphisms $\sigma: A \to A$, and $\tau: B \to B$ such that the composition $(\tau \circ \varphi \circ \sigma) = \tilde{\varphi}: A \to B$ has a diagonal matrix, then M is isomorphic to a product of cyclic groups.

Remark: A finite cyclic group of order r in this product corresponds to a diagonal entry equal to r in the matrix $[\tilde{\varphi}]$. Hence the number of non zero diagonal entries equals the number of finite cyclic factors in the product, and if the diagonal entries divide each other, then the product decomposition is in standard form.

§2) Diagonalizing an integral matrix, with application to homomorphisms of free abelian groups

By the previous corollary, the next proposition will imply the decomposition theorem.

Proposition: If $A = \mathbf{Z}^n$, $B = \mathbf{Z}^m$, and $\varphi: A \to B$ is a homomorphism, then there exist isomorphisms $\sigma: A \to A$, and $\tau: B \to B$ such that the composition $(\tau \circ \varphi \circ \sigma) = \tilde{\varphi}: A \to B$ has a diagonal matrix.

The proof is by means of "elementary row and column operations", similar to "Gaussian elimination" for solving linear equations in several variables. Since we want our operations to be reversible over $\mathbf{Z}$, we restrict to multiplication by units in operation 3) below.

Elementary row and column operations on integer matrices

1) Interchange two rows (or two columns) of the matrix.
2) Add to one row (or column) an integer multiple of another.
3) Multiply through a row (or column) by a unit, i.e. by 1 or -1.

The previous proposition follows from the next two lemmas.

Lemma 1: If $A = \mathbf{Z}^n$, $B = \mathbf{Z}^m$, and $\varphi: A \to B$ is a homomorphism with matrix $|\varphi|$, and if $|\tilde{\varphi}|$ is any matrix obtained from $|\varphi|$ by a sequence of row and column operations, then there are isomorphisms $\sigma: A \to A$, and $\tau: B \to B$ such that the composition $(\tau \circ \varphi \circ \sigma) = \tilde{\varphi}: A \to B$ has matrix equal to $|\tilde{\varphi}|$.

Lemma 2: A matrix of integers can always be transformed into a diagonal matrix by some sequence of row and column operations.

Idea for proof of lemma 1: This is the same as in matrix theory courses, i.e. each row and column operation can be done by multiplying by an appropriate "elementary" matrix, the matrix of the corresponding isomorphism. For completeness, we will define matrix multiplication below and remark that matrix multiplication corresponds to composition of homomorphisms.

An algorithm for the proof of lemma 2:

The idea is to try to make the upper left entry in the matrix as small as possible. If all entries in the matrix are zero, there is nothing to do: it is already diagonal. If some matrix entry is non zero, bring the smallest non zero entry (in absolute value) to the top

left position by interchanging its row with the first row and its column with the first column.

Now the first entry in the first column is the smallest non zero entry in the matrix. Call it a. We next try to make it even smaller, and make all other entries in the first column equal to zero as follows. Consider the second entry in the first column, say it is b. If b is divisible by a, then subtract an appropriate multiple of the first row from the second row replacing b by zero, and proceed to the third entry in the first column. If b is not divisible by a, subtracting an appropriate multiple of the first row from the second row replaces b by a positive integer smaller than a in absolute value. Then interchanging first and second rows makes the first entry in the first row smaller than it was. Now repeat this procedure. Eventually we get the gcd of a and b as the first entry in the first column. Then the first entry divides the second and the next step replaces the second entry by zero.

Then we proceed to the third entry and repeat the process. At the end of this procedure every entry in the first column is zero except the first entry, which has been getting continually smaller in absolute value.

Next we carry out the same process on the first row, replacing all entries except the first by zero, and always making the first entry smaller. Unfortunately this process may destroy what we did to the first column, i.e. those entries may not be zero any longer. So we return to the first column and repeat the procedure again, making all entries zero again except the first, which continues to get smaller. Then we go back again to the first row, whose entries may no longer be zero and repeat the procedure again, etc.

Since we are continually making the first entry in the first column smaller, as long as that entry fails to divide some entry in the first row or the first column, and since that first entry cannot get smaller forever, eventually entry (1,1) is non zero and all other entries in the first row and the first column are zero.

Now consider the submatrix of rows and columns after the first, and repeat the whole procedure. If it is not identically zero, eventually

we get a non zero entry in the (2,2) position, and all other entries in the second row and second column are zero. Note that operations performed on these later rows and columns do not change any of the entries in the first row and column, because in those positions these operations only add zeroes to zeroes. Continuing to the next submatrix, and so on, we eventually get a diagonal matrix, with the non zero diagonal entries occurring before the zero entries.

Now to get the matrix in standard form, add all the columns after the first to the first column, putting all the diagonal entries into the first column. Then repeat the original procedure on the first column, getting a diagonal matrix again but with the new (1,1) entry equal to the gcd of all the diagonal entries. Now repeat this procedure on the submatrix of rows and columns after the first, getting the gcd of the remaining diagonal entries as new (2,2) entry. Eventually the matrix is diagonal and each diagonal entry divides the remaining ones.

Finally, if necessary, we can multiply through selected rows by -1 until all non zero entries are positive. Now the matrix is diagonal, the diagonal entries successively divide each other, and the non zero entries on the diagonal are all positive. **QED lemma 2.**

Remark: There is no need to follow this algorithm in reducing any particular matrix. In particular cases you should carry out whatever operations look best and most efficient to you. The algorithm given always works, hence proves it can be done, and gives at least one way to program a computer to diagonalize a matrix In practice it should always be easier than this.

An example is probably preferable to the previous discussion.

$$\begin{bmatrix} 7 & 12 & 11 \\ 5 & 2 & -3 \\ 10 & 14 & 10 \\ -6 & -12 & -12 \end{bmatrix} \approx \begin{bmatrix} 7 & 12 & 11 \\ 5 & 2 & -3 \\ 10 & 14 & 10 \\ 1 & 0 & -1 \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & -1 \\ 5 & 2 & -3 \\ 10 & 14 & 10 \\ 7 & 12 & 11 \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \\ 0 & 14 & 20 \\ 0 & 12 & 18 \end{bmatrix}$$

$$\approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 6 \\ 0 & 0 & 6 \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{bmatrix}.$$

Consequently, if $H \subset (\mathbf{Z}\times\mathbf{Z}\times\mathbf{Z}\times\mathbf{Z})$ is the subgroup generated by the columns of the first matrix above, the quotient $M = (\mathbf{Z}\times\mathbf{Z}\times\mathbf{Z}\times\mathbf{Z})/H$ is isomorphic to $\{0\}\times\mathbf{Z}_2\times\mathbf{Z}_6\times\mathbf{Z}$. In particular rank(M) = 1.
QED lemma 2.

Background for proof of lemma 1: Review of Dot products, Matrix products and sums

We define the dot product of two elements $(a_1,\ldots,a_n)$, $(b_1,\ldots,b_n)$ of $\mathbf{Z}^n$ to be $(a_1,\ldots,a_n)\cdot(b_1,\ldots,b_n) = \Sigma a_i b_i$. Then we define the product AB of two integer matrices to be the matrix whose entry in the ith row and jth column equals the dot product of the ith row of A with the jth row of B. This product is defined iff the rows of A are the same length as the columns of B, i.e. iff the number of columns of A equals the number of rows of B. Thus if AB is defined, BA may still not be defined. For example, if A is 2×5 and B is 5×4, the product AB is defined and is a 2×4 matrix, but the product BA is not defined. Almost any example shows even if both AB and BA are defined, they need not be equal.

We can add two matrices iff they have the same dimensions, by adding corresponding entries. We can multiply a matrix by an integer, by multiplying every entry by that integer. Matrix multiplication commutes with integer multiplication and is distributive over matrix addition as long as all operations are defined. The "zero matrix", with all zero entries, is an additive identity for matrices of its size, and the "identity matrix", a square matrix with 1's on the diagonal and zeroes elsewhere, is the multiplicative identity for matrices whose product with it is defined.

The product of matrices is associative when defined. An example you work out yourself is probably the best way to be convinced of this, but for what it's worth: if $A = (a_{ij})$, $B = (b_{jk})$, and $C = (c_{kl})$ then the (i,k) entry of AB is $\Sigma_j (a_{ij}b_{jk})$, so the (i,l) entry of (AB)C is $\Sigma_k(\Sigma_j(a_{ij}b_{jk})c_{kl})$. Similarly, since the (j,l) entry of BC is $\Sigma_k(b_{jk}c_{kl})$, the (i,l) entry of A(BC) is $\Sigma_j a_{ij}(\Sigma_k b_{jk}c_{kl})$. Interchanging order of

summation, the (i,l) entry of $(AB)C = \Sigma_k\Sigma_j(a_{ij}b_{jk}c_{kl}) = \Sigma_j\Sigma_k(a_{ij}b_{jk}c_{kl})$ = the (i,l) entry of A(BC). Hence (AB)C = A(BC).

The homomorphism associated to a matrix

If A is an m×n matrix, then A defines a homomorphism $\varphi:\mathbf{Z}^n\to\mathbf{Z}^m$, where for any element $x = (x_1,\ldots,x_n)$ of $\mathbf{Z}^n$, $\varphi(x) = Ax$ with x written as a column vector on the right side of the matrix A. Thus Ker(φ) = the subgroup of vectors x in $\mathbf{Z}^n$ whose dot product with every row of A is zero. It is easy to see that $\varphi(e_j)$ = the jth column of A. Hence $\varphi(x) = \Sigma_j x_j\varphi(e_j)$ = the linear combination of the columns of A formed with the coefficients from x. Consequently, Im(φ) = the subgroup of $\mathbf{Z}^m$ generated by the columns of A. If A,B are matrices whose product AB is defined, since for every x in $\mathbf{Z}^n$ (AB)x = A(Bx) by associativity, if φ is the homomorphism defined by A and ψ that defined by B, the product AB is the matrix of the composition φ∘ψ. Since the identity matrix is a multiplicative identity, it is the matrix of the identity homomorphism. Thus a homomorphism φ is an isomorphism iff its matrix has a multiplicative inverse, and then that inverse matrix is the matrix of the inverse isomorphism φ^{-1}.

If $M \cong \mathbf{Z}^m$, $N \cong \mathbf{Z}^n$, it follows from our discussion that the additive group Hom(N,M) ≅ {additive group of all m×n integer matrices} ≅ $\mathbf{Z}^{mn}$, the free abelian group on mn elements. Moreover, the non commutative (if $m \geq 2$) ring End(M) = Hom(M,M) ≅ {non commutative (for $m \geq 2$) ring of all m×m matrices}. The additive group of this ring is isomorphic to the free abelian group on m^2 elements, but the ring structure is not that of the product ring $\Pi\mathbf{Z}$ with m^2 factors, since that ring is commutative. If we denote the invertible m×m matrices by $GL_m(\mathbf{Z})$, then Aut(M) = End(M)* = the unit group of the ring End(M), is isomorphic to the multiplicative group $GL_m(\mathbf{Z})$, ("general linear group over Z of order m").

proof of lemma 1:

We need the following facts: If A is an m×n matrix, and [τ] is the matrix obtained from the m×m identity matrix by performing any row operation, then the product [τ]A is the matrix obtained by performing the same row operation on A. If [σ] is the matrix

obtained from the n×n identity matrix by performing any column operation, then A[σ] is the matrix obtained by performing the same column operation on A. We could prove these statements without too much trouble, but again it is much more enlightening for you to carry out some examples.
Now we can prove lemma 1, since performing a sequence of row and column operations on A is equivalent to multiplying A by a sequence of matrices as follows $[\tau_r][.....][\tau_1]A[\sigma_1][....][\sigma_s]$. Since each row or column operation is inverted by another row or column operation, each matrix $[\tau_i]$, $[\sigma_j]$, is invertible, hence is the matrix of an isomorphism τ_i, and σ_j respectively. Since a composition of isomorphisms is an isomorphism, the compositions $\tau = (\tau_r \circ ..... \circ \tau_1)$ and $\sigma = (\sigma_1 \circ .... \circ \sigma_s)$, are isomorphisms Hence if A is the matrix of a homomorphism $\varphi: \mathbf{Z}^n \to \mathbf{Z}^m$, then any matrix obtained from φ by row and column operations is the matrix of a composition of form $(\tau \circ \varphi \circ \sigma)$ where τ, σ are isomorphisms. QED for **lemma 1**.

This finally proves the diagonalization proposition, and thus the fundamental theorem of finitely generated abelian groups.
QED Theorem.

Exercise #125) If K is the subgroup of $\mathbf{Z}^5$ generated by the elements (3,0,15,3,15), (6,6,-6,12,0), and (24,6,-6,30,0), find the standard decompositon of the quotient group $M = \mathbf{Z}^5/K$. What is the rank of M? the invariant factors?

Remarks on Rank and Independence:
If M is any abelian group with *m* generators, and $\mathbf{Z}^m \to M$ a surjection, the kernel K of $\mathbf{Z}^m \to M$ is always free of rank $n \leq m$. Hence the map $\varphi: \mathbf{Z}^n \to \mathbf{Z}^m$ such that $M \cong \mathbf{Z}^m/\mathrm{Im}(\varphi)$ can always be assumed to be injective, with rank(K) = n = number columns of the matrix $[\varphi]$. Since $\mathrm{Ker}(\varphi) = \{0\}$, it follows from the definition of matrix multiplication that 0 is the only element of $\mathbf{Z}^n$ which dots to zero with every row of $[\varphi]$. Since by exercise #113), the kernel is unchanged, up to isomorphism, by row and column operations, this remains true after diagonalizing $[\varphi]$. Hence all n diagonal entries of the diagonalized matrix $[\tilde{\varphi}]$ are non zero, so there are n finite cyclic factors in the decomposition of M (some possibly zero), and (m-n) infinite cyclic factors. In particular the rank of M equals (m-n).

Exercise #126) Prove that rank is additive in the sense that: If $M \cong A/B$ where A is any fin. gen abelian group and $B \subset A$ a subgroup, then rank(M) = rank(A) - rank(B).

Definition: A subset S of an abelian group M is *(linearly) independent* iff whenever a finite linear combination $\Sigma \alpha_i x_i = 0$, with x_i in S, α_i in $\mathbf{Z}$, then all $\alpha_i = 0$. Hence a subset S is *(linearly) dependent iff* there exists some finite expression $\Sigma \alpha_i x_i = 0$, with x_i in S, α_i in $\mathbf{Z}$, and some $\alpha_i \neq 0$.

Exercise #127) Prove:
(i) An independent set cannot contain any elements of finite order.
(ii) If M has an independent generating set S, M is free on S.
(iii) The rank of a finitely generated abelian group M equals the largest integer n such that M contains an independent subset of cardinality n.

Definition: An independent generating set $S \subset M$ in an abelian group M is called a "basis" of M. (M may or may not *have* a basis.) A "minimal set of generators" is one which no longer generates when any element is removed, and an "maximal independent set" is one which is no longer independent when any element is adjoined.

The following assertions are analogs of those in a theorem about finite dimensional vector spaces in section §15 of the 843 notes:
Consider these assertions:
If $M \cong \mathbf{Z}^n$, then:
i) More than n elements of M are always dependent.
ii) Fewer than n elements cannot generate M.
iii) Every basis of M has exactly n elements.
iv) Every set of n independent elements in M also generates M.
v) Every set of n generators of M is also independent.
vi) A maximal independent subset of M has exactly n elements.
vii) Every minimal set of generators of M has exactly n elements.
viii) Every independent subset of M is contained in a basis.
ix) Every generating set for M contains a basis.
x) A subgroup of M cannot have rank greater than n.
xi) A proper subgroup of M has rank less than n.
xii) An isomorphism $f: M \to N$ carries a basis of M to a basis of N, hence N is also free of rank n

Exercise #128) Which of the 12 assertions above are true?

Exercise #129) Give counterexamples to the assertions above which are false.

Exercise #130) Prove the assertions above which are true.

§3) Diagonalizing a matrix over a Euclidean domain, with application to homomorphisms of "free R-modules".

It should be clear that the process given (in the proof of lemma 2 in the previous section) for diagonalizing a metrix of integers, consisted essentially of using row/column operations to replace two integers by their gcd, i.e. to carry out the Euclidean algorithm. This means it can be repeated in almost the same way in any ring admitting such an algorithm, i.e. in any Euclidean domain, such as the polynomial ring k[X] over a field. To exploit this observation, in this section we define matrices over any ring R, and investigate what they represent. An m×n matrix over any ring R is of course a rectangular array of elements of R with m rows and n columns. We assume as usual that R is commutative with identity.

Definition: $\mathrm{Mat}_{m\times n}(R)$ is the abelian group of m×n matrices over R, where two matrices are added by adding corresponding entries. If we define multiplication of R-matrices exactly as for **Z**-matrices [the (i,j) entry of AB is the dot product of the ith row of A with the jth column of B, an element of R], then $\mathrm{Mat}_{n\times n}(R)$ is a ring (non commutative for $n\geq 2$) whose group of units we denote by $GL_n(R)$.

Note: As additive group, $\mathrm{Mat}_{m\times n}(R) \cong R^{mn}$, the product of mn copies of the abelian additive group underlying R.

To generalize the diagonalization process to matrices over a Euclidean domain R, we need to modify our row/column operations so they allow the Euclidean algorithm in R to be carried out in R. The following definitions make sense over any ring R.

Elementary row/column operations on an R-matrix:

1) Interchange two rows (or two columns) of the matrix.
2) Add to one row (or column) an R- multiple of another.
3) Multiply through a row (or column) by a unit of R.

Diagonalizing a matrix: To deduce the diagonalization theorem, it suffices to require R to possess a Euclidean algorithm.

Lemma: If R is a Euclidean domain, any R-matrix can be diagonalized by elementary row and column operations.
proof: The algorithm given for the proof of lemma 2 in section §2 works here, but let's think it through again
If some entries are non zero, bring (one of) the smallest (in Euclidean "size") non zero entry to the top left position by interchanging its row and column with the first row and column. Call it a.

If the second entry in the first column, say b, is divisible by a, subtract an appropriate multiple of the first row from the second row replacing b by zero, and proceed to the third entry in the first column. If b is not divisible by a, subtracting an appropriate (R-) multiple of the first row from the second row replaces b by an element of R smaller than a in size. Interchanging first and second rows makes the first entry in the first row smaller than it was. Repeating this procedure, eventually makes the gcd of a and b the first entry in the first column, and the next step replaces the second entry by zero.

Proceeding to the third entry, we repeat the process, eventually replacing every entry in the first column by zero except the first, which has been getting continually smaller in size.

Doing the same process to the first row, replaces all entries except the first by zero, always making the first entry smaller, but possibly introducing non zero entries in the first column. Returning to the first column and repeating the procedure again, makes all those entries zero again except the first, which continues to get smaller. We repeat the procedure again, on the first row, then the first column again, etc.... Since entry (1,1) cannot get smaller forever, eventually it is non zero and all other entries in the first row and the first column are zero.

By induction, the submatrix of rows and columns after the first can be diagonalized by row/column operations, and these do not disturb the first row/column. The same procedure as before puts the matrix in standard form, i.e. the matrix is diagonal and each

diagonal entry divides the remaining ones.

The last step in the previous argument of making all diagonal entries positive integers makes no sense here. Hence the diagonal entries are not unique here, but only up to multiplication by units of R. QED.

Interpreting the result: Now what good does this do us, i.e. what does it mean? Eg. what does an m×n matrix of entries in a ring R represent? Since integer matrices represent homomorphisms $\mathbf{Z}^n \to \mathbf{Z}^m$ of free abelian groups, surely an R-matrix gives some sort of homomorphism of the groups $R^n \to R^m$. If we define it the same way as before, i.e. an m×n matrix A over R defines the function $\varphi(x) = A \cdot x$, where x is written as a column matrix, then we do get a group homomorphism, but in fact not a general one. I.e. with this definition φ has the property that $\varphi(\alpha x) = \alpha\varphi(x)$ for every x in R^n and every α in R, where $\alpha(x_1,\dots,x_n) = (\alpha x_1,\dots,\alpha x_n)$. The point is that our group R^n is a special one, with more than just the usual additive group structure: it also admits multiplication by elements of R, and the homomorphism defined by a matrix not only commutes with addition, but also with multiplication by R.

Let us make this precise. First define "R-multiplication" on a group.
Definition: Given an abelian group (M,+), and a (commutative with identity) ring R, an **R-module** structure on M is a pairing $R \times M \to M$ with the usual properties of multiplication. That is, if we write simply αm or $\alpha \cdot m$ for the image of (α,m) in M, then for all α,β in R, and all m,n in M, we have $(\alpha\beta)m = \alpha(\beta m)$, $(\alpha+\beta)m = \alpha m + \beta m$, $\alpha(n+m) = \alpha m + \alpha n$, and $1 \cdot m = m$.
An abelian group M with such a structure is called an R-module.

Remarks: i) The third property in the definition above says the elements of R become group homomorphisms $M \to M$, via multiplication, i.e. each r in R defines a homomorphism $r \cdot M \to M$, by $r(x) = rx$. Thus an R-module structure on M defines a function $R \to \mathrm{Hom}(M,M) = \mathrm{End}(M)$. The first and second properties above say this function is a ring homomorphism $R \to \mathrm{End}(M)$ into the non commutative ring End(M) (where multiplication is composition).
ii) Conversely, if $\sigma: R \to \mathrm{End}(M)$ is any ring homomorphism onto a

(commutative since R is) subring of End(M), we get an R-module structure on M by setting $rx = \sigma(r)(x)$. The point is that there is one smallest and one largest ring naturally acting on M, the largest being the non commutative ring End(M), and the smallest being the subring generated by id_M which gives the natural $\mathbb{Z}$-action on M, via the unique map $\mathbb{Z} \to End(M)$ taking 1 to id_M. Our hope is there may be lots of interesting commutative rings R in between these two, whose action can tell us more about M.
iii) Note that a group M may not admit a structure of R-module for every ring R. For instance as we have noted before, if M is a finite abelian group and R is an infinite field, then there is no ring map $R \to End(M)$ [A ring map with field as source is injective, but if R is infinite and End(M) is finite, no such map is possible.] Precisely, M has an R-module structure iff the ring End(M) contains a subring isomorphic to a quotient of R by an ideal. The only ring R such that End(M) always has this property is $R = \mathbb{Z}$.
iv) Since every R -module structure on M is given by a ring map $R \to End(M)$, there is an associated kernel of this map, the so called "annihilator ideal" for the R-module. I.e. $R \supset ann(M) = \{$those r in R such that $rx = 0$ for all x in M$\}$. Thus the structure map factors via an injective ring map $R/ann(M) \to End(M)$ displaying M also as an R/ann(M) - module, where R/ann(M) is isomorphic to a commutative subring of End(M).
v) Viewing R-multiplication as a ring map $R \to End(M)$ is analogous to our earlier viewpoint that an action of a group G on a set S is equivalent to a group homomorphism $G \to Bij(S)$. Since our ring R is commutative, here we may ignore the problem of distinguishing between a "left" or "right" module structure.
vi) If R is a field, an R-module is nothing but a vector space over R. If $R=\mathbb{Z}$, an R-module is precisely an abelian group.
vii) The word "module" was apparently introduced by number theorists at the end of the 19th century to denote the ring of integers in a number field, an important example of a finitely generated abelian group.

Remark: We have defined a new concept analogous to an abelian group. Our first job is to examine this analogy, in order to take advantage of what we already know about abelian groups. So we will rethink a number of fundamental constructions on abelian groups, to see how to modify them in the present setting.

As always, the first necessity, in order to compare two R-modules, is to define "R-homomorphisms".
Definition: If M, N are two R modules, a group homomorphism $\varphi: N \to M$ such that $\varphi(rx) = r\varphi(x)$ for every r in R, and every x in N, is called an R-module homomorphism, or **R-homomorphism**, or R-module map, or simply a homomorphism or even just a map.
[I.e. if everybody knows the objects are R-modules, it may often be assumed that "map" means R-module map.]

The set of all R-module maps $N \to M$ is denoted $\mathbf{Hom_R(N,M)}$. It is itself an R-module where $(\varphi+\psi): N \to M$ is defined as usual by adding values, i.e. $(\varphi+\psi)(x) = \varphi(x)+\psi(x)$, and where $(r\varphi)(x) = r\cdot(\varphi(x))$
When N = M, $\mathrm{Hom_R(M,M)} = \mathbf{End_R(M)}$ is a non commutative ring whose multiplication is composition. Since R is commutative, the ring map $R \to \mathrm{End(M)}$ actually has image in $\mathrm{End_R(M)}$ and in fact in the "center" of this ring, i.e. multiplication by elements of R commutes (by definition) with R-endomomorphisms of M. The ring structure on $\mathrm{End_R(M)}$ together with the ring map of R into its center, says $\mathrm{End_R(M)}$ is more than just an R-module, it is an "**R-algebra**". [If S is a commutative ring, an R-algebra structure on S is just a ring map $R \to S$; in particular this makes S an R - module.]

The most important definition, which tells us when two R-modules are essentially "alike" is the following:
Definition: An **R-isomorphism**, or simply isomorphism, of two R-modules N,M, is an R-module map $\varphi: N \to M$ with an R-inverse, i.e. one admitting another R-module map $\psi: M \to N$ such that $\varphi\circ\psi = \mathrm{id_M}$, and $\psi\circ\varphi = \mathrm{id_N}$.

Notation: We denote the set (possibly empty) of R-isomorphisms $N \to M$ by $\mathbf{Isom_R(N,M)}$, and the non abelian group $\mathrm{Isom_R(M,M)} = \mathbf{Aut_R(M)}$, (always non empty since it contains $\mathrm{id_M}$).

Remark: The crucial definition of isomorphism is anticlimactic, since we learned when discussing categories that the definition of an isomorphism (of any type) is always an admissible map φ such that some other admissible map ψ exists of that same type, in the other direction, such that $\varphi\circ\psi$ and $\psi\circ\varphi$ are both identity maps. The next result makes proving the existence of such inverses a little easier:

Exercise #131) An R-module map is an isomorphism iff it is bijective.

Remarks: i) Not all group maps are R-module maps, i.e. the inclusion $\mathrm{Hom}_R(N,M) \subset \mathrm{Hom}_{\mathbb{Z}}(N,M)$ is usually proper. For instance, $\mathbb{R}$ is an abelian group and an infinite dimensional vector space over $\mathbb{Q}$. Thus we can choose, by Zorn's Lemma, a $\mathbb{Q}$-basis for $\mathbb{R}$. Then we can define a $\mathbb{Q}$-linear (additive in particular) map $\varphi:\mathbb{R} \to \mathbb{R}$ sending each of the infinitely many basis elements anywhere we like. In particular, since $\{1,\pi\}$ is a $\mathbb{Q}$-linearly independent set, hence contained in a basis, there need be no relation at all between $\varphi(1)$ and $\varphi(\pi)$. Now $\mathbb{R}$ is also an $\mathbb{R}$-module, but the only $\mathbb{R}$-module maps $\varphi:\mathbb{R} \to \mathbb{R}$ are given by multiplication by a fixed real number α. I.e. if $\varphi(1)=\alpha$, then $\varphi(\pi) = \pi\varphi(1) = \pi\alpha$. More generally, if φ is an $\mathbb{R}$-module map, then $\varphi(x) = x\varphi(1)$, for all x in $\mathbb{R}$.
A simpler example: the quotient ring $R = \mathbb{Z}[X]/(X^2)$. As abelian group, $R \cong \mathbb{Z}\times\mathbb{Z}$, where (n,m) corresponds to $[n+mX]$. Thus the map $\varphi:R \to R$ taking (n,m) to $(2n,3m)$ i.e. such that $\varphi([n+mX]) = [2n+3X]$, is a group map but not an R-module map, since that would require eg., $\varphi([X][n+mX]) = [X]\varphi([n+mX]$. But the l.h.s. equals $\varphi([nX] = [3nX]$, while the r.h.s. equals $[2nX]$, a contradiction.
ii) Since R-modules generalize vector spaces, abelian groups, and ideals, to appreciate statements about R-modules, it helps to translate them into each of these cases.

At last we can say **what m×n matrices over R are for**: they represent R-homomorphisms $R^n \to R^m$. Note that $\mathrm{Mat}_{m\times n}(R)$ has an R-module structure, where we multiply a matrix A by an element α of R by multiplying every entry of A by α.
Proposition: There is a natural isomorphism of R-modules $\mathrm{Mat}_{m\times n}(R) \cong \mathrm{Hom}_R(R^n,R^m)$.
proof: There is a standard basis for R^n consisting of the vectors $\{e_j\}$, where $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0)$, etc....., and R^n is free on this set in the sense that an R- homomorphism $R^n \to X$ is determined by where we send these basis elements; moreover we may send them anywhere we please. Thus an R-homomorphisms out of R^n is determined by an n-tuple of elements of the target module, in this case an n-tuple of elements of R^m, i.e. an n-tuple of column vectors of length m with entries in R. This is precisely an $m\times n$ matrix over

R. This correspondence preserves addition and R-multiplication. To be precise, define a map $\Theta:\mathrm{Hom}_R(R^n,R^m)\to \mathrm{Mat}_{m\times n}(R)$, by $\Theta(\varphi) = [\varphi(e_1)\ \varphi(e_2)\ldots\ldots\varphi(e_n)]$. I.e. $\Theta(\varphi)$ is the matrix whose jth column is the vector $\varphi(e_j)$. Then $\Theta(\varphi+\psi) = [(\varphi+\psi)(e_1)\ (\varphi+\psi)(e_2)\ldots\ldots(\varphi+\psi)(e_n)] = [\varphi(e_1)\ \varphi(e_2)\ldots\ldots\varphi(e_n)] + [\psi(e_1)\ \psi(e_2)\ldots\ldots\psi(e_n)] = \Theta(\varphi) + \Theta(\psi)$. Also $\Theta(r\varphi) = [(r\varphi)(e_1)\ (r\varphi)(e_2)\ldots\ldots(r\varphi)(e_n)] = r\cdot[\varphi(e_1)\ \varphi(e_2)\ldots\ldots\varphi(e_n)] = r\cdot\Theta(\varphi)$. Thus Θ is an R-homomorphism.
There is also has an inverse map Ω as follows:
$\Omega:\mathrm{Mat}_{m\times n}(R) \to \mathrm{Hom}_R(R^n,R^m)$, where $\Omega([v_1\ldots\ldots v_n]) = \varphi_v$, and $\varphi_v(\alpha_1,\ldots,\alpha_n) = \Sigma\alpha_j v_j = v\cdot\alpha$ = the matrix product of the matrix v with the *column* vector α.
To see these are inverses, $(\Theta\circ\Omega)([v_1\ldots\ldots v_n])$ is the matrix whose jth column is $\varphi_v(e_j) = v\cdot e_j = v_j$. And $(\Omega\circ\Theta)(\varphi)$ is the map whose value at $\alpha = (\alpha_1,\ldots,\alpha_n)$ is $\varphi_{\varphi(e)} = \varphi(e)\cdot\alpha = \Sigma\alpha_j\varphi(e_j) = \varphi(\alpha)$. Thus $(\Omega\circ\Theta)(\varphi) = \varphi$. QED.

Remark: We repeat, if $\varphi:R^n\to R^m$ is a homomorphism, its standard matrix is the m×n matrix whose jth *column* is the vector $\varphi(e_j)$.

Corollary(of the proof): For any R-module M, $\mathrm{Hom}_R(R^n,M) \cong M^n$.
proof: Any $\varphi:R^n\to M$ is uniquely determined by where it sends the standard basis $\{e_j\}$ and we can send those elements anywhere we wish, so φ is equivalent data to the n-tuple $(\varphi(e_1),\ldots,\varphi(e_n))$ in M^n. This shows there is a bijection of sets, and an argument analogous to that above shows this is an R-module isomorphism, where the R-module structure on the Cartesian product M^n is defined componentwise, [see §4, (i), below]. QED.

Remark: Multiplying matrices over R is still associative, by the proof given above in §2 over **Z** ["interchange order of summation", which depended on associativity of multiplication and commutativity of addition in R]. Consequently, matrix multiplication still corresponds to composition of homomorphisms, hence gives an isomorphism $\mathrm{Mat}_n(R) \cong \mathrm{End}(M)$ as rings.

Now we can restate the diagonalization result for matrices over Euclidean domains, in terms of homomorphisms, as we did over **Z**.

Proposition: If R is any Euclidean domain, $A = R^n$, $B = R^m$, and $\varphi: A \to B$ is a homomorphism, there exist isomorphisms $\sigma: A \to A$, and $\tau: B \to B$ such that the composition $(\tau \circ \varphi \circ \sigma) = \tilde{\varphi}: A \to B$ has a diagonal matrix, in which each diagonal entry divides the next.
proof: Just as in §2 above (over **Z**), this proposition follows from the diagonalization result above and the next lemma:

Lemma: If R is any ring, $A = R^n$, $B = R^m$, and $\varphi: A \to B$ is a homomorphism with matrix $[\varphi]$, and if $[\tilde{\varphi}]$ is any matrix obtained from $[\varphi]$ by a sequence of row and column operations, then there are isomorphisms $\sigma: A \to A$, and $\tau: B \to B$ such that the composition $(\tau \circ \varphi \circ \sigma) = \tilde{\varphi}: A \to B$ has matrix equal to $[\tilde{\varphi}]$.
proof: The proof of this is word for word the same as the proof of lemma 1 in the previous section (replacing **Z** by R). In particular, the isomorphisms σ, τ are again compositions of "elementary" isomorphisms, where an elementary isomorphism is one whose matrix is the matrix of an elementary row/column operation.
QED. for both Lemma and Prop.

Remarks: i) Note that row and column operations are equivalent to elementary isomorphisms, over any ring, but that compositions of elementary isomorphisms do not necessarily suffice to diagonalize a matrix except over a Euclidean ring.
ii) To prove that every finitely generated module over any Euclidean domain, is isomorphic to a product of cyclic modules, we can just repeat the proof given over **Z**, but first we need to give R-module versions of the definitions and standard properties of the "nouns" in that proof, such as: submodule, cyclic module, product module, quotient module, image module, first isomorphism theorem. We also need to prove submodules of finite free modules are finitely generated. These are entirely analogous to the proofs for abelian groups, but since we want to be confident about them, and use them in other contexts as well, we spend the next two sections discussing these foundations, and then complete the decomposition theorem in section §6 below

Digression on "naturality": When we proved the equivalence of matrices and homomorphisms above, we did not say what we meant by "natural", although you may feel that it reflects the predictable nature of the isomorphisms we gave in the proof. In fact

:t has a more precise "functorial" meaning. As Auslander put it, "it :means when you change the objects, you get maps, and everything :ommutes" More precisely, suppose you have a ring map $\sigma: R \to S$, :.e. suppose you "change the ring". Then by applying σ to every :ntry in an R-matrix A, you get an S matrix $\sigma(A)$, hence a map $\tilde{\sigma}: \mathrm{Mat}_{m\times n}(R) \to \mathrm{Mat}_{m\times n}(S)$. Applying σ to every entry of a column :ector also gives a map $\sigma: R^m \to S^m$, and we can define $\tilde{\sigma}: \mathrm{Hom}_R(R^n, R^m) \to \mathrm{Hom}_S(S^n, S^m)$, where for φ in the source, and $\alpha = (\alpha_1, \ldots, \alpha_m)$ we define $\sigma(\varphi)(\alpha) = \Sigma \alpha_j(\sigma\circ\varphi)(e_j)$, where $\{e_j\}$ is the standard basis of R^n. Then naturality means that the two composite maps $\mathrm{Mat}_{m\times n}(R) \to \mathrm{Hom}_R(R^n, R^m) \to \mathrm{Hom}_S(S^n, S^m)$, and $\mathrm{Mat}_{m\times n}(R) \to \mathrm{Mat}_{m\times n}(S) \to \mathrm{Hom}_S(S^n, S^m)$, are the same, i.e. $\tilde{\sigma} \circ \Omega_R = \Omega_S \circ \tilde{\sigma}$.
Examples of change of rings are $\mathbf{Z} \to \mathbf{Z}_n$, $\mathbf{Z} \subset \mathbf{Q}$, $\mathbf{Q} \subset \mathbf{R}$, $\mathbf{R} \subset \mathbf{C}$, $k \subset k[X]$.

In the next section we show how to produce lots of R-modules, give their basic properties, and some important occurrence of modules in nature, such as vector fields, and rings of integers in number fields.

§4) Examples and constructions of R-modules

Just as for vector spaces, the most fundamental concept is that of representing elements of modules as linear combinations of other elements, including possible uniqueness of such representations.

Definition: Given any collection of elements $S = \{x_j\}$ of an R-module M, a **linear combination** (or R-linear combination) of elements of S is a finite expression of form $\Sigma \alpha_j x_j$, with the α_j in R and x_j in S.

Definition: A subset $\{x_j\} \subset M$ is **linearly independent** iff whenever $\Sigma \alpha_j x_j = 0$, with α_j in R, then all $\alpha_j = 0$.]

1) Submodules: A subgroup of M which is closed under R-multiplication is called an R-submodule of M. The subset of M consisting of 0 and all linear combinations of elements of S is a submodule called the **R-submodule generated by S.** (Hence the empty set $\varnothing$ generates {0}.) Thus S is a **generating set for M iff** every non zero element of M is a linear combination of elements of S. A module is called **finitely generated** iff it can be generated by a finite set S.

M can be generated by a subset of m elements iff there is a surjective homomorphism $R^m \to M$, for some positive integer m. A submodule $(a) = Ra \subset M$, generated by one element a in M, is called "cyclic" or "**principal**". For any R-homomorphism $\varphi: M \to N$, the subgroup $M \supset \ker(\varphi) = \{x \text{ in } M : \varphi(x) = 0\}$ is an R submodule of the source, since if $\varphi(x) = 0$, then also $\varphi(rx) = r\varphi(x) = r \cdot 0 = 0$, so if x is in $\ker(\varphi)$ then rx is in $\ker(\varphi)$. For any R-homomorphism $\varphi: M \to N$, the subgroup $N \supset \mathrm{Im}(\varphi) = \{y \text{ in } N \,.\, y = \varphi(x), \text{ for some } x \text{ in } M\}$ is an R submodule of the target, since if $y = \varphi(x)$, then $ry = \varphi(rx)$, so if y is in $\mathrm{Im}(\varphi)$ then ry is in $\mathrm{Im}(\varphi)$ too.
A subset $I \subset R$ is an R-submodule iff it is an R-ideal. The subset $\{(a,a)$ for all a in $R\}$ is the R-submodule of R^2 generated by $\{(1,1)\}$. This might be called the "diagonal" submodule of R^2. If $I \subset R$ is an ideal, and M is an R-module, the subset $M \supset IM = \{$linear combinations $\Sigma \alpha_j x_j$ with α_j in I, x_j in $M\}$ is a submodule of M. It is the submodule generated by all products $\{\alpha x, \alpha$ in I, x in $M\}$.

2) Finite product modules and (finite rank) free modules.
If $M_1,\ldots,M_s$ are R-modules, the Cartesian product group $M = \prod M_j$ is naturally an R-module where $\alpha(a_1, \ldots, a_n) = (\alpha a_1,\ldots,\alpha a_n)$ for α in R, and a_j in M_j. In particular, since R is a module over itself, the product group R^n has a natural structure of R-module. For $n = 1$, R is thus a module over itself.
If I,J are ideals of R, $I \times J$ is a submodule of R^2. If I,J are principal, generated by a,b respectively, then $I \times J$ is generated by $\{(a,0), (0,b)\}$. More generally, if $N_1 \subset M_1$, and $N_2 \subset M_2$ are submodules, then $N_1 \times N_2$ is a submodule of $M_1 \times M_2$.
The projection $\pi_i: \prod_j M_j \to M_i$, on the ith factor, i.e. $\pi_i(a_1,\ldots,a_s) = a_i$, is an R-module map. The projection $\pi_i: R^n \to R$ has matrix $[0\ldots0\ 1\ 0\ldots0]$, where the "1" is in the ith position. Note, for $x = (x_1,\ldots,x_s)$ in $\prod_j M_j$, $\pi_i(x) = x_i$, so $x = (\pi_1(x),\ldots,\pi_s(x)) = (x_1,\ldots,x_s)$.
The injection $\sigma_i: M_i \to \prod_j M_j$, of the ith factor, $\sigma_i(a_i) = (0,\ldots,0,a_i,0,\ldots0)$, is an R-module map. The injection $\sigma_i: R \to R^n$ has matrix $\begin{bmatrix} 0 \\ . \\ 1 \\ . \\ 0 \end{bmatrix}$ where the "1" is in the ith position.

Definition: An R-module which is isomorphic to R^n is called "free" of rank n. [See below for proof that rank is well defined.]

For example, $Mat_{m\times n}(R) \cong R^{mn}$ is a free R-module of rank mn, and thus so is $Hom_R(R^n,R^m)$, by a result in the previous section. If I,J are principal ideals in a domain R, then $I\times J \subset R^2$ is a free submodule: i.e. if I = (a), J =(b), respectively, then $\varphi: R^2 \to I\times J$ defined by $\varphi(1,0) = (a,0)$, $\varphi(0,1) = (0,b)$, is an isomorphism.

Proposition: The following are equivalent for an R-module M:
(i) M is free, of finite rank.
(ii) M is isomorphic to R^n for some n.
(iii) M has a finite subset $\{x_j\}$ such that each element of M has a unique expression as an R-linear combination of elements of $\{x_j\}$.
(iv) M has a finite subset $\{x_j\}$ such that for any R-module X, every function $\{x_j\} \to X$ extends uniquely to an R-module map $M \to X$.
(v) M can be generated by a finite linearly independent subset.
proof: Useful exercise.

Definition: A "basis" or "free basis" for an R module M is a linearly independent generating set. Thus M is free iff it has a basis.

Caution: Some texts refer to a "basis" of an ideal in a ring R, when they mean only a "generating set". Indeed, a non empty subset $S \subset R$ is linearly independent iff S consists of one non zero-divisor. (Two elements $\{a,b\} \subset R$ must be dependent since ba-ab=0.) Hence only *principal* ideals can have *free* bases.

7) Modules of homomorphisms.
As observed above, for any two R-modules N, M, the set of R-homomorphisms $Hom_R(N,M)$ is naturally an R-module. The particular example where M = R, i.e. $Hom_R(N,R)$, is called the "dual module of N", and denoted N^*. Note: this module may not be as big as it looks: $Hom_{\mathbb{Z}}((\mathbb{Z}_{10})^{100}, \mathbb{Z}) = \{0\}$. [Why?] If we put the ring in the first variable of Hom, we do not get anything new, in the following sense:

Lemma: There is a natural isomorphism $\mathrm{Hom}_R(R,M) \cong M$. More precisely, "the functor $\mathrm{Hom}_R(R,\cdot)$ is naturally equivalent to the identity functor" (on R-modules).
proof: The isomorphism follows from our earlier result that $\mathrm{Hom}_R(R^n,M) \cong M^n$, but naturality needs to be checked. Auslander used to summarize naturality simply as: "when you change the objects you get maps, and everything commutes". Here there are two objects, R and M, but we have stated the result as dealing with functors of M, so M is the object and the functors $\mathrm{Hom}_R(R,\cdot)$ and $\mathrm{Id}_{\mathcal{M}}$ are thought of as acting on the category $\mathcal{M}$ of R-modules M. I.e. the Hom functor sends $M \mapsto \mathrm{Hom}_R(M)$, and the identity functor sends $M \mapsto M$. Since these are functors we must also say what they do to maps. For $\mathrm{Id}_{\mathcal{M}}$ this is trivial, since it obviously sends $\varphi: N \to M$ to $\varphi: N \to M$, thus it certainly sends identities to identities and compositions to compositions, so it is a functor.

The Hom(R,·) functor on the other hand must send a map $\varphi: N \to M$ to some map $\varphi_*: \mathrm{Hom}_R(R,N) \to \mathrm{Hom}_R(R,M)$. The only natural choice is "composition with φ". Thus given f in $\mathrm{Hom}_R(R,N)$, define $\varphi_*(f) = \varphi \circ f$. Then $(\psi\circ\varphi)_*(f) = (\psi\circ\varphi)\circ f = \psi\circ(\varphi\circ f) = \psi_*(\varphi_*(f)) = (\psi_*\circ\varphi_*)(f)$. So $(\psi\circ\varphi)_* = (\psi_*\circ\varphi_*)$, hence the assignment does preserve compositions. Since $(\mathrm{id}_M)_*(f) = (\mathrm{id}_M)\circ f = f$, thus $(\mathrm{id}_M)_* = \mathrm{id}_{\mathrm{Hom}(R,M)}$, so the assignment also sends identities to identities. Naturality of the isomorphism Θ: $\mathrm{Hom}_R(R,M) \cong M$, means that, given a map φ $N \to M$, if we use the isomorphisms Θ_M and Θ_N to equate $\mathrm{Hom}_R(R,M)$ with M, and $\mathrm{Hom}_R(R,N)$ with N, then φ must become equated with φ_*. In other words, if we recall the isomorphism Θ_N: $\mathrm{Hom}_R(R,N) \to N$, to be given by $\Theta(f) = f(1)$, then the compositions $\mathrm{Hom}_R(R,N) \to \mathrm{Hom}_R(R,M) \to M$, and $\mathrm{Hom}_R(R,N) \to N \to M$, must be equal; i.e. $\Theta_M \circ \varphi_* = \varphi \circ \Theta_N$. So let f belong to $\mathrm{Hom}_R(R,N)$. Then $(\Theta_M \circ \varphi_*)(f) = \Theta_M(\varphi\circ f) = (\varphi\circ\varphi)(1) = \varphi(f(1))$. On the other hand $(\varphi \circ \Theta_N)(f) = \varphi(\Theta_N(f)) = \varphi(f(1))$. QED.

The following notion of maps "of pairs", i.e of maps which respect submodules, has interest for the next section on maps of quotient modules: if M and X are any modules with submodules $N \subset M$, $Y \subset X$, the set $\mathrm{Hom}_R((M,N),(X,Y))$ of "R-homomorphism of the pairs" (M,N), (X,Y), is defined as those R-maps $\varphi: M \to X$ such that $\varphi(N) \subset Y$. This

subset $\mathrm{Hom}_R((M,N), (X,Y)) \subset \mathrm{Hom}_R(M,X)$ is closed under addition and R-multiplication, hence is a submodule. In particular, $\mathrm{Hom}_R(M,X) \supset \mathrm{Hom}_R((M,N), (X,0))$ = those R-maps $\varphi: M \to X$ such that $N \subset \ker(\varphi)$, is a submodule.

4) Quotient modules. If $N \subset M$ is an R-submodule of an R-module, then the quotient group M/N has a natural R-module structure, where $r[x] = [rx]$. This is well defined since if $[x]=[y]$, then $x-y$ is in N, so $r(x-y) = rx - ry$ is in N, and thus $[rx] = [ry]$. This definition says precisely that the canonical group map $M \to M/N$ is also an R homomorphism.
For example, up to isomorphism, the only quotient modules of R are those of form R/I where $I \subset R$ is an ideal. If $(a_1,\dots,a_m)$ is an element of R^m, we can construct from it two submodules and quotient modules: either $R^m \supset N$ = the principal submodule $R\cdot(a_1,\dots,a_m) = \{(ra_1,\dots,ra_m)$, for all r in R$\}$, with quotient R^m/N, or the product $\Pi_j (a_j)$ of principal submodules of R, with quotient $R^m/(\Pi_j (a_j)) \cong \Pi_j (R/(a_j))$. More generally, if $N = \Pi_j I_j$, for any R-ideals $I_1,\dots,I_m$, then $R^m/N \cong \Pi_j (R/I_j)$.

Remark: An R-module M is cyclic iff M is isomorphic to one of form R/I for some ideal $I \subset R$. Note, if $R = \mathbf{Z}[X]$, then R is cyclic as an R-module, but not cyclic as a group, i.e. not as a $\mathbf{Z}$-module. In fact R is not even finitely generated as a $\mathbf{Z}$-module.

If $I \subset R$ is an ideal and M an R-module, and $M \supset IM$ = the submodule generated by products αx with α in I, x in M, then the quotient M/IM is an R-module. But M/IM is also naturally an R/I module, since multiplication by elements of I annihilates M/IM, so the ring map $R \to \mathrm{End}(M/IM)$ factors via a ring map $R/I \to \mathrm{End}(M/IM)$.

5) An example from geometry and physics: vector fields.
Let $S = \{p=(x,y,z): x^2+y^2+z^2 = 1\}$, be the unit sphere in $\mathbb{R}^3$, and let $T \subset S \times \mathbb{R}^3$ be the subset of pairs $\{(p,v)$, s.t. $p \cdot v = 0\}$. T is called the "tangent bundle" of S, since each pair (p,v) in T consists of a point p of S and a vector v parallel to the plane tangent to S at p. [If you want to visualize (p,v) as an actual geometric tangent vector to S, you should think of it as representing the vector whose foot is at p and whose head is at p+v; i.e. our vector v is the unique vector

whose foot is at the origin, and which is parallel to the actual geometric tangent vector. For example, (p,v) = ((0,0,1), (1,0,0) should be visualized as representing the geometric tangent vector with foot at the north pole (0,0,1) of S and head at (0,0,1)+(1,0,0) = (1,0,1).] A pair (p,v) in T is called a tangent vector to S at p. A "section" of the tangent bundle, or "vector field" on S, is any right inverse $\sigma: S \to T$ to the "first" projection $\pi_1: S \times \mathbb{R}^3 \supset T \to S$, where $\pi_1(p,v) = p$. I.e. $\sigma: S \to T$ is a vector field iff for each p in S, $\sigma(p)$ is a tangent vector at p. If we denote by $T_p = \pi_1^{-1}(p) = \{p\} \times \{v \text{ in } \mathbb{R}^3 : p \cdot v = 0\}$ the set of tangent vectors to S at p, then a section of T is a function $\sigma: S \to T$ such that for all p in S, $\sigma(p)$ is in T_p. A continuous vector field is a section $\sigma: S \to T$ whose composition $(\pi_2 \circ \sigma): S \to S \times \mathbb{R}^3 \to \mathbb{R}^3$ with the "second" projection $\pi_2: T \to \mathbb{R}^3$ (where $\pi_2(p,v) = v$), is continuous. If $\mathcal{V}$ is the set of all continuous vector fields on S, then $\mathcal{V}$ is an abelian group under the following operation: if $\sigma(p) = (p,v)$ and $\tau(p) = (p,w)$, then $\sigma(p)+\tau(p) = (p,v+w)$, i.e. we do the addition in the second variable. Now if $\mathcal{C}$ = {all continuous real valued functions $f: S \to \mathbb{R}$ on the sphere}, $\mathcal{C}$ is a commutative ring with identity, and $\mathcal{V}$ is a $\mathcal{C}$ module, where if $\sigma(p) = (p,v)$ then $(f \cdot \sigma)(p) = (p, f(p) \cdot v)$.
If $H \subset S$ is a hemisphere, $\mathcal{V}_H$ the set of continuous vector fields on H, and $\mathcal{C}_H$ the ring of continuous functions on H, then $\mathcal{V}_H \cong \mathcal{C}_H \times \mathcal{C}_H$ is a free $\mathcal{C}_H$ - module. In general $\mathcal{V} \cong \mathcal{C} \times \mathcal{C}$ iff there exist two sections σ, τ in $\mathcal{V}$ such that for every p in S, $\sigma(p)$ and $\tau(p)$ are linearly independent tangent vectors. There is a very interesting theorem in topology that says for every σ in $\mathcal{V}$, there exists some p in S such that $\sigma(p) = 0$; i.e. there is not an everywhere non zero, continuous vector field on the "two-sphere". Thus in fact not even one section σ exists with $\sigma(p)$ independent for all p, hence $\mathcal{V}$ is not isomorphic to $\mathcal{C} \times \mathcal{C}$ I.e. $\mathcal{V}$ is not free over $\mathcal{C}$. Since $\mathcal{V}$ is free over every hemisphere in S we can say $\mathcal{V}$ is at least "locally free", however.

Exercise #132) Prove: (i) If there exist two sections σ, τ in $\mathcal{V}$ such that for every p in S, $\sigma(p)$ and $\tau(p)$ are linearly independent tangent vectors, then $\mathcal{V} \cong \mathcal{C} \times \mathcal{C}$. (ii) If $H = \{(x,y,z): x^2+y^2+z^2 = 1, \text{ and } z > 0\}$ is the "northern" hemisphere of S, then $\mathcal{V}_H \cong \mathcal{C}_H \times \mathcal{C}_H$, with notation as in the previous example.

6) An example from number theory: rings of integers.
The three types of rings we have studied, Euclidean domains, p.i.d.'s,

and u.f.d.'s, are important, but unfortunately the basic rings of number theory, the rings of "integers", usually do not belong to any of these types. There is another weaker property that characterizes them, the property of "integral closure". If $R \subset S$ are rings, an element x of S is "integral over R" iff x satisfies a monic polynomial over R. A ring R is "integrally closed" iff the only elements of the quotient field of R which are integral over R, are elements of R itself. Since the only rational roots of a monic integral polynomial are integers, by the "rational root theorem", $\mathbf{Z}$ is integrally closed. The same proof used for rational root theorem shows that every ufd is integrally closed.

If $\mathbf{Q} \subset F$ is a finite extension field, F is called a "number field" and the subset $O \subset F$ of elements of F integral over $\mathbf{Z}$ is called the ring of "integers in F". In particular O is a ring, in fact an integrally closed noetherian domain in which every proper prime ideal is maximal, (according to the "fundamental theorem of algebraic number theory"). [It is useful to think of the length of a maximal chain of proper prime ideals as the "dimension" of a ring, so these are "one dimensional" integrally closed rings.] Since $\mathbf{Z} \subset O$, O is a $\mathbf{Z}$-algebra, but in fact O is a finitely generated free $\mathbf{Z}$-module of rank $n = [F:\mathbf{Q}]$. For example, the ring of Gaussian integers $\mathbf{Q}[i] \supset \mathbf{Z}[i] \cong \mathbf{Z}^2$. If $\mathbf{Q} \subset K \subset F$ are two number fields with integers $R \subset S$, then S is also finitely generated as R module, but not necessarily free. More generally, if R is an integrally closed noetherian domain with quotient field K, $K \subset F$ a finite extension field, and $R \subset S$ the integral closure of R in F, then S is a noetherian integrally closed domain, and a finitely generated R module, but not necessarily free unless R is a p.i.d. Recall the word "module" was introduced to describe this example, by number theorists in the 19th century.

§5) How to define homomorphisms on products, quotients. Before we prove various R-modules are isomorphic to products of cyclic modules, we need to know how to define maps in and out of products and quotients. The facts are entirely analogous to ones about abelian groups, as we have observed in several examples above, but we will try to discuss them here more systematically. First we generalize our earlier result on maps out of free modules to show how to define **maps out of a finite product.**

Proposition: If $M_1,\ldots,M_s$, and X, are R-modules, there is a natural isomorphism $\mathrm{Hom}_R(\Pi_i M_i,X) \cong \Pi_i \mathrm{Hom}_R(M_i,X)$.
In words: defining a map out of a finite product is equivalent to giving one map out of each factor.
proof: To find this isomorphism, we follow Auslander's advice: *first* just find natural maps in both directions, *then* check whether they are mutually inverse. The point is that since the isomorphism is natural, the only maps you can think of are surely the right ones.

We define a map Θ: $\mathrm{Hom}_R(\Pi_i M_i,X) \to \Pi_i \mathrm{Hom}_R(M_i,X)$, where if $\varphi:\Pi_i M_i \to X$, then $\Theta(\varphi) = (\varphi_1,\ldots,\varphi_s)$ with $\varphi_i:M_i \to X$ defined by $\varphi_i(a_i) = \varphi(0,\ldots,a_i,\ldots\ldots 0)$. I.e. just "restrict" φ to each factor separately.

The inverse is $\Omega:\Pi_i \mathrm{Hom}_R(M_i,X) \to \mathrm{Hom}_R(\Pi_i M_i,X)$, where $\Omega(\varphi_1,\ldots,\varphi_s) = \varphi:\Pi_i M_i \to X$, with $\varphi(a_1,\ldots,a_s) = \Sigma\varphi_i(a_i)$. I.e. just add the separate images of the coordinates. [This needs a finite product to work.]

Check: Ω and Θ are both R-module maps.
Eg If $(\varphi_1,\ldots,\varphi_s)$, $(\psi_1,\ldots,\psi_s)$, are in $\Pi_i \mathrm{Hom}_R(M_i,X)$, and $a = (a_1,\ldots,a_s)$ in $\Pi_i M_i$, then $\Omega((\varphi_1,\ldots,\varphi_s)+(\psi_1,\ldots,\psi_s))(a_1,\ldots,a_s) = \Sigma(\varphi_i+\psi_i)(a_i) = \Sigma\varphi_i(a_i)+\Sigma(\psi_i(a_i) = \Omega(\varphi_1,\ldots,\varphi_s)(a)+\Omega(\psi_1,\ldots,\psi_s)(a)$. Hence Ω is additive. Since $\Omega(r\varphi_1,\ldots,r\varphi_s)(a) = \Sigma(r\varphi_i)(a_i) = r\Sigma\varphi_i(a_i) = r\cdot\Omega(\varphi_1,\ldots,\varphi_s)(a)$, Ω is R-linear. We omit the check for Θ.

Check: Ω and Θ are mutually inverse.
If $(\varphi_1,\ldots,\varphi_s)$ is in $\Pi_i \mathrm{Hom}_R(M_i,X)$, then $(\Theta\circ\Omega)(\varphi_1,\ldots,\varphi_s) = (\psi_1,\ldots,\psi_s)$, where $\psi_i(a_i) = (\Omega(\varphi_1,\ldots,\varphi_s))(0,\ldots,0,a_i,0,\ldots,0) =$
$\varphi_1(0)+\ldots+\varphi_{i-1}(0)+\varphi_i(a_i)+\varphi_{i+1}(0)+\ldots+\varphi_s(0) = \varphi_i(a_i)$.
Thus $(\Theta\circ\Omega)(\varphi_1,\ldots,\varphi_s) = (\varphi_1,\ldots,\varphi_s)$.

If φ is in $\mathrm{Hom}_R(\Pi_i M_i,X)$, $\Theta(\varphi) = (\varphi_1,\ldots,\varphi_s)$, $a = (a_1,\ldots,a_s)$ in $\Pi_i M_i$, then $(\Omega\circ\Theta)(\varphi)(a) = (\Omega(\Theta(\varphi)))(a) = (\Omega(\varphi_1,\ldots,\varphi_s))(a) = \Sigma_i \varphi_i(a_i) =$
$\Sigma_i \varphi(0,\ldots,a_i,\ldots\ldots 0) = \varphi(\Sigma_i (0,\ldots,a_i,\ldots\ldots 0)) = \varphi(a)$. Thus $(\Omega\circ\Theta)(\varphi) = \varphi$.
QED.

Exercise #133) Check the map Θ in the previous proof is an R-module map.

Remark: The previous proposition is false for infinite products.

The next lemma describes **maps into products.**
Proposition: If If $M_1,\dots,M_s$, and X, are R-modules, then $\mathrm{Hom}_R(X,\prod_i M_i) \cong \prod_i \mathrm{Hom}_R(X,M_i)$.
In words: defining a map into a product is equivalent to defining one map into each factor.
proof: First we just look for some obvious maps. Given a homomorphism into a product, we can get a map into each factor by projecting on the factor.
Hence if we let $\pi_i:\prod_i M_i \to M_i$ be the projection on the ith factor, we can define a map in one direction, as follows: $\Theta:\mathrm{Hom}_R(X,\prod_i M_i) \to \prod_i \mathrm{Hom}_R(X,M_i)$ is defined by $\Theta(\varphi) = (\pi_1\circ\varphi,\dots,\pi_s\circ\varphi)$.

To go the other way, given a map of X into every factor, we get a map into the product by viewing those as the component maps. I.e. this is just as in calculus where a "vector valued" function is given by two, three, or n, real valued "component" functions. So we get the map $\Omega:\prod_i \mathrm{Hom}_R(X,M_i) \to \mathrm{Hom}_R(X,\prod_i M_i)$ defined by $\Omega(\varphi_1,\dots,\varphi_s)(x) = (\varphi_1(x),\dots,\varphi_s(x))$.
We omit verification that Θ, Ω are R-module maps. [See Ex.123.]

Check Θ, Ω are mutually inverse:
$(\Theta\circ\Omega)(\varphi_1,\dots,\varphi_s) = \Theta(\Omega(\varphi_1,\dots,\varphi_s)) = (\pi_1\circ\Omega(\varphi_1,\dots,\varphi_s),\dots,\pi_s\circ\Omega(\varphi_1,\dots,\varphi_s))$ $= (\psi_1,\dots,\psi_s)$, where $\psi_i(x) = (\pi_i\circ\Omega(\varphi_1,\dots,\varphi_s))(x) = \pi_i(\varphi_1(x),\dots,\varphi_s(x)) = \varphi_i(x)$. Thus $(\Theta\circ\Omega)(\varphi_1,\dots,\varphi_s) = (\psi_1,\dots,\psi_s) = (\varphi_1,\dots,\varphi_s)$.

$(\Omega\circ\Theta)(\varphi)(x) = \Omega(\pi_1\circ\varphi,\dots,\pi_s\circ\varphi)(x) = ((\pi_1\circ\varphi)(x),\dots,(\pi_s\circ\varphi)(x)) = (\pi_1(\varphi(x)),\dots,\pi_s(\varphi(x))) = \varphi(x)$. QED.

Exercise #134) In the special case of two factors, show the maps Θ, Ω in the previous proof between $\mathrm{Hom}_R(X,M\times N)$ and $\mathrm{Hom}_R(X,M)\times\mathrm{Hom}_R(X,N)$, are R-module maps.

Remark: The previous proposition is <u>still true</u> for infinite products.

Next we characterize **maps out of quotients.**
Lemma: If $N\subset M$ is a submodule, and $\varphi:M\to X$ a module map such that $\varphi(N) = \{0\}$, [i.e. any homomorphism of "pairs" $\mathrm{Hom}_R((M,N),(X,0))$], then φ induces a unique homomorphism $\tilde{\varphi}:M/N\to X$ such that $\tilde{\varphi}([x]) = \varphi(x)$ for each [x] in M/N.

proof: Since, as an abelian group, M/N is the quotient group of M by the subgroup N, we know this map is a well defined group map, so we just need to check it is also a module map. But for r in R, $\tilde{\varphi}(r[x]) = \tilde{\varphi}([rx]) = \varphi(rx) = r\varphi(x) = r\tilde{\varphi}([x])$. QED.

Corollary: There is a natural injection $\mathrm{Hom}_R(M/N,X) \bigotimes \mathrm{Hom}_R(M,X)$, such that $\mathrm{Hom}_R(M/N, X) \cong \mathrm{Hom}_R((M,N), (X,0)) \subset \mathrm{Hom}_R(M,X)$.

There is no natural characterization of maps *into* a quotient, i.e. although every map $X \to M$ induces by composition a map $X \to M/N$, there are maps $X \to M/N$ which are not of this type, i.e. which do not "lift" to M. [For instance, the identity map $\mathbf{Z}_n \to \mathbf{Z}_n$ does not factor through $\mathbf{Z}_n \to \mathbf{Z} \to \mathbf{Z}_n$.] Also different maps into M can become the same map into M/N. [Both the maps $x \mapsto nx$ and $x \mapsto 0$, from $\mathbf{Z} \to \mathbf{Z}$, become zero as maps $\mathbf{Z} \to \mathbf{Z}_n$.] Thus although there is a natural map from $\mathrm{Hom}_R(X,M) \to \mathrm{Hom}_R(X,M/N)$, it is neither injective nor surjective.

We do have analogs of the usual isomorphism theorems:
Lemma: (i) If $\varphi: M \to N$ is a surjective R module map, then the induced map $\tilde{\varphi}: M/\ker(\varphi) \to N$ is an isomorphism.
(ii) If $B,C \subset M$ are submodules, then the map $B \to (B+C)/C$ induced by inclusion $B \to B+C$, has kernel $B \cap C$, hence induces $B/(B \cap C) \cong (B+C)/C$.
(iii) If $C \subset B \subset M$ are submodules, then $\varphi: M \to M/B$ implies $C \subset \ker(\varphi)$, hence induces $M/C \to M/B$ with kernel B/C, hence $(M/C)/(B/C) \cong M/B$.
proof: Useful exercise.

Now let's prove that "rank" of a finite free module is well defined.
Theorem: If R is a ring and $R^n \cong R^m$, as R modules, then $n = m$.
proof: We know this result for vector spaces, i.e. if R is a field, since this says dimension is well defined. I.e. a vector space isomorphism must carry a basis to a basis, so this follows from the theorem that all bases of a given vector space have the same cardinality. Let's try to reduce the present statement to the case of a vector space over a field. The only natural way to pass from an arbitrary (always commutative with identity) ring R to a field, is to mod out by a maximal ideal, so let $I \subset R$ be maximal. If $R^n \cong R^m$, then $I \cdot R^n \cong I \cdot R^m$, whence $(R/I)^n \cong R^n/I^n \cong R^n/I \cdot R^n \cong R^m/I \cdot R^m \cong (R/I)^m$. This proves $(R/I)^n \cong (R/I)^m$, as R modules, but since the ideal $I \subset R$

kills both of these modules, they are also R/I modules, and these maps are R/I isomorphisms. Since R/I is a field, $(R/I)^n$ and $(R/I)^m$ are isomorphic R/I vector spaces, of dimensions n, m respectively. Since a vector space isomorphism carries a basis into a basis, they must have the same dimension, i.e. n = m. QED.

Exact sequences:
In our proof of the decomposition theorem for finitely generated abelian groups, we have found it useful on occasion to construct compositions of maps $\varphi\circ\psi$ such that $\ker(\varphi) = \mathrm{Im}(\psi)$. This phenomenon is so commonly encountered that a special name has been given to it: "exactness". We examine this concept briefly.

Definition: A sequence of R-homomorphisms $\ldots\to A\to B\to C\to \ldots$ is called "exact at B" iff the image of the incoming map equals the kernel of the outgoing map, i.e. if $\mathrm{Im}(A\to B) = \ker(B\to C)$. A sequence is simply called "exact" iff it is exact everywhere.

Exercise #135) If A,B,C are R modules, prove:
(i) $0\to A\to B$ is exact at A, iff $A\to B$ is injective.
(ii) $A\to B\to 0$ is exact at B, iff $A\to B$ is surjective.
(iii) Two maps $\varphi:A\to B$ and $\psi:B\to C$, yield an exact sequence $0\to A\to B\to C\to 0$, iff $\varphi:A\to B$ is injective and ψ induces an isomorphism $B/\varphi(A)\cong C$.
(iv) If $0\to A_1\to A_2\to \ldots\ldots\to A_n\to 0$ is an exact sequence of finitely generated abelian groups, prove the "alternating sum" of the ranks is zero, i.e. $[\mathrm{rk}(A_1)-\mathrm{rk}(A_2)+\mathrm{rk}(A_3)-\ldots\ldots \pm \mathrm{rk}(A_n)] = 0$

§6) Decomposing finitely generated modules over p.i.d.'s.

We will prove that a finitely generated modules over a p.i.d. is isomorphic to a product of cyclic modules, the best generalization possible for the fundamental theorem of finite abelian groups. In general R denotes a p.i.d. in this section unless we say otherwise.

Theorem: If R is a p.i.d. and M is a finitely generated R module, there exist unique integers $s,t \geq 0$, and a unique nested sequence of proper ideals $R\supset(d_1)\supset(d_2)\supset\ldots\ldots\supset(d_t)$, such that
$M \cong R^s \times \prod_{i=1\ldots t} R/(d_i)$.
Terminology: The integer s is called the "rank" or "R-rank" of M.

proof of Uniqueness: We need to generalize the concept of elements of finite order so we can distinguish the "free" factors isomorphic to R from the "torsion" factors isomorphic to R/(d).

Definition: An element x of M, a module over any ring R, is called a "torsion element" iff there is a non zero element r of R such that rx = 0. If R⊃ann(x) = {r in R: rx = 0} is the "annihilator ideal of x, then x is a torsion element iff ann(x) ≠ {0}. We denote ann(M) = $\cap_x$ ann(x), the ideal of elements annihilating everything in M.

Exercise #136) (i) If R is a domain, and M is an R-module, prove the subset of torsion elements of M forms a submodule, t(M)⊂M.
(ii) Give an example of an R module M whose subset of torsion elements is not a submodule, (and R is not a domain).
(iii) Prove, if R is a domain, and $M = R^s \times \Pi_{i=1...t}\, R/(d_i)$, where for each i, $(d_i) \neq R$, then $t(M) = \{0\} \times \Pi_{i=1...t}\, R/(d_i)$.
(iv) If R is a domain, and φ:M→N is an isomorphism, φ(t(M)) = t(N).

Now assume $\varphi: M = R^s \times \Pi_{i=1...t}\, R/(d_i) \to R^u \times \Pi_{j=1...v}\, R/(e_j) = N$, is an isomorphism, where again R is a p.i.d. By exercise #125), φ restricts to an isomorphism $t(M) = \{0\} \times \Pi_{i=1...t}\, R/(d_i) \cong \{0\} \times \Pi_{j=1...v}\, R/(e_j) = t(N)$. Thus φ induces an isomorphism of the quotients M/t(M) ≅ N/t(N), hence $R^s \cong R^v$. Then s = v, by the last theorem of the previous section, (that the rank of a finite free module is always well defined). This proves the uniqueness of the "free part" of M, i.e. of the integer s in the theorem.

To prove the uniqueness of the torsion part, we use the ideas from the proof in the finite case, i.e. analogs of the Sylow p-subgroups.
Definition: If r ≠ 0 is an element of R, let M(r) = {x in M: rx = 0}, the "r-torsion" of M, (a submodule), and let $M_r = \cup_{n>0}\, M(r^n)$, be the set of elements annihilated by powers of r, also a submodule.

Definition: If R is a p.i.d., and x an element of an R module, the "order of x" means any generator of the ideal ann(x). So r is the order of x iff ann(x) = (r)⊂R.

The "order" is no longer an integer but an element of R, and is only defined up to multiplication by units of R. If p is a prime element of

the p.i.d. R, and M is an R module, then the p-torsion submodule $M(p)$ = {set of elements of order p}, and M_p = {set of elements whose order is a power of p}. Thus M_p generalizes the Sylow p-subgroup of a finite abelian group (when p is a prime integer).

Exercise #137) Assume R is a p.i.d. and p is a prime element.
If $M \cong A \times B$, prove:
(i) $M(p^s) \cong A(p^s) \times B(p^s)$ for all s;
(ii) $M_p \cong A_p \times B_p$.
If d is a non unit of R, $M = R/(d)$, prove
(iii) if $p^r | d$ then $M(p^r) \cong R/(p^r)$;
(iv) if $p^r | d$ but $p^{r+1} \nmid d$, then $M_p = M(p^r) \cong R/(p^r)$.

To continue the proof of uniqueness of the torsion part of a fin. gen. module over a p.i.d., let $A = \prod_{i=1...t} R/(d_i) \cong \prod_{j=1...v} R/(e_j) = B$, where $d_i | d_{i+1}$, and $e_j | e_{j+1}$. We want to show that $t = v$ and that $(d_i) = (e_i)$ for all $1 \le i \le t$. It is easy to show the smallest ideals equal.

Exercise #138) With the assumptions of the previous paragraph, prove: (i) $A_p \neq \{0\}$ iff $p | d_t$, and $B_p \neq \{0\}$ iff $p | e_v$;
(ii) these ideals are equal: $(d_t) = \mathrm{ann}(A) = \mathrm{ann}(B) = (e_v)$;
(iii) $A \cong B$ implies $A_p \cong B_p$.

To show the other ideals (d_i) are (e_i) are equal, we will show the generators d_i , e_i, have the same prime factors. So let p be a factor of $d_t = u \cdot e_v$, assume $p^{r_i} | d_i$ but $p^{r_i+1} \nmid d_i$, and $p^{s_j} | e_j$ but $p^{s_j+1} \nmid e_j$. Then by the previous two exercises, since $A \cong B$ we have $\prod_{i=1...t} R/(p^{r_i}) \cong A_p \cong B_p \cong \prod_{j=1...v} R/(p^{s_j})$, where $r_i, s_j \ge 0$. Moreover, since $d_i | d_{i+1}$, and $e_j | e_{j+1}$, we have $p^{r_i} | p^{r_{i+1}}$, and $p^{s_j} | p^{s_{j+1}}$, i.e. $r_i \le r_{i+1}$, and $s_j \le s_{j+1}$.

Lemma: If p is a prime element of R (a p.i.d.), if $\prod_{i=1...\alpha} R/(p^{r_i}) \cong \prod_{j=1...\beta} R/(p^{s_j})$, and $1 \le r_i \le r_{i+1}$, $1 \le s_j \le s_{j+1}$, then $\alpha = \beta$, and for all $1 \le i \le \alpha$, $r_i = s_i$.
proof: First we prove $\alpha = \beta$. If we denote $C = \prod_{i=1...\alpha} R/(p^{r_i})$, and $D = \prod_{j=1...\beta} R/(p^{s_j})$, then by the previous two exercises, $C \cong D$ implies $\prod_{i=1...\alpha} R/(p) \cong C(p) \cong D(p) \cong \prod_{j=1...\beta} R/(p)$. Since in a p.i.d.

every prime ideal is maximal, R/(p) is a field and thus C(p) is a vector space over R/(p) of dimension α, while D(p) is an R/(p) vector space of dimension β. By invariance of dimension of vector spaces, $\alpha = \beta$. Now we can finish by induction on r_α. I.e. if $r_\alpha = 1$, then all $r_i = 1$, and C = C(p), so D = D(p), hence all $s_j = 1$. QED for this case. If we have proven the result for $r_\alpha < n$, and if we have $r_\alpha = n$, then consider $\Pi_{i=1...\alpha} R/(p^{r_i-1}) \cong C/C(p) \cong D/D(p) \cong \Pi_{i=1...\alpha} R/(p^{s_i-1})$. By induction on the largest exponent occurring in the left module, we have that on both sides the same number of non zero exponents occur, and those exponents are equal. Hence also the same number of zero exponents occur. Thus $r_i = s_j$ for all i. **QED lemma.**

Again let $A = \Pi_{i=1...t} R/(d_i) \cong \Pi_{j=1...v} R/(e_j) = B$, where $d_i | d_{i+1}$, and $e_j | e_{j+1}$. We want to show that $t = v$ and that $(d_i) = (e_i)$ for $1 \leq i \leq t$. We know $(d_t) = (e_v)$, and since $d_i | d_t$ for all i, and $e_j | e_v$ for all j, the prime factors of the $\{d_i\}$ are exactly the prime factors of d_t = the prime factors of e_v = the prime factors occurring in the $\{e_j\}$. We have just shown that every prime p occurs in the same number of factors of A as of B, and with the same exponents. Since a prime factor of d_1 occurs in all t factors of C, it also occurs in t factors of D. Hence $t \leq v$. By symmetry $v \leq t$, so $t = v$. Since every prime occurs with the same exponents in the factors of C and of D, the elements d_i, e_i must be associates. Hence $(d_i) = (e_i)$ for all i.
QED for uniqueness of decomposition.

Proof of existence:
If M is a finitely generated module over R, a p.i.d., then there is an R module surjection $\sigma: R^m \to M$, for some m. Then $M \cong R^m/\ker(\sigma)$, where $\ker(\sigma)$ is a submodule of R^m. To represent this submodule as the image of another such map we need the submodule $\ker(\sigma)$ to be finitely generated.

Lemma: Every submodule of R^m is free of rank $\leq m$, if R is a p.i.d.
proof: The proof is the same as for **Z**, by induction. I.e. if $m = 1$, a submodule $A \subset R$ is an ideal hence principal, hence either $= (0)$ and thus free of rank zero, or $= (x)$ for some $x \neq 0$ in R. In the second case, $R \to (x)$ taking r to rx is an isomorphism and (x) is free of rank one. For $m > 1$, consider the projection map $R^m \to R$, where

$\pi(x_1,\ldots,x_m) = x_m$, and restrict it to a submodule $A \subset R^m$. Then we have $\pi: A \to R$, and $\mathrm{Im}(\pi) \subset R$ is an ideal hence free of rank ≤ 1. Then $\ker(\pi)$ is a submodule of $R^{m-1} \times \{0\} \cong R^{m-1}$, so by induction $\ker(\pi)$ is free of rank $\leq m-1$. If $\mathrm{Im}(\pi) = \{0\}$, then $A \cong \ker(\pi)$ is free of rank $\leq m-1 < m$ and we are done. If $\mathrm{Im}(\pi) \neq 0$, then $\mathrm{Im}(\pi) \cong R$, $\ker(\pi) \cong R^t$ for some $t \leq m-1$, and thus we have an exact sequence $0 \to R^t \to A \to R \to 0$, where $t \leq m-1$. We claim the sequence splits, so that $A \cong R^t \times R \cong R^{t+1}$. By the next exercise, it suffices to show the surjective map $\pi: A \to R$ has a right inverse. To define one, let a be any element of A with $\pi(a) = 1$ in R. Then there is a unique R-module map $\psi: R \to A$ with $\psi(1) = a$, namely $\psi(r) = ra$. Since $\pi(a) = 1$, $\pi(ra) = r$, hence $\pi(\psi(r)) = r$ for all r in R, and ψ is right inverse to π. Then $A \cong R^{t+1}$, where $t+1 \leq m$. QED.

Exercise #139) Prove: for any ring R, a surjective R-module map $f: A \to B$ splits, i.e. $A \cong B \times \ker(f)$, if f has a right inverse.

Now we have $M \cong R^m/\ker(\sigma)$, and $\ker(\sigma) \cong R^n$ for some $n \leq m$. Thus there is an injective R-module map $\varphi: R^n \to R^m$ with $M \cong R^m/\mathrm{Im}(\varphi)$, and φ is given by an $m \times n$ matrix $[\varphi]$ with entries in R. The following proposition would complete the proof:

Proposition: If R is any p.i.d., $A = R^n$, $B = R^m$, and $\varphi: A \to B$ is a homomorphism, there exist isomorphisms $\sigma: A \to A$, and $\tau: B \to B$ such that the composition $(\tau \circ \varphi \circ \sigma) = \tilde{\varphi}: A \to B$ has a diagonal matrix, in which each diagonal entry divides the next.

Assuming this proposition, we have $M \cong R^m/\mathrm{Im}(\tilde{\varphi})$, where $\mathrm{Im}(\tilde{\varphi})$ is generated by $(d_1,0,\ldots,0)$, $(0,d_2,0,\ldots,0)$, $(0,0,d_3,0,\ldots 0)$,..., $(0,\ldots,0,d_n,0,\ldots,0)$. Thus $M \cong$
$(R \times R \times \ldots \times R)/(d_1R \times d_2R \times \ldots \times d_nR \times \{0\} \times \ldots \times \{0\}) \cong$
$R/(d_1) \times R/(d_2) \times \ldots \times R/(d_n) \times R \times \ldots \times R$, where each d_i divides d_{i+1}, and there are $m-n$ factors of R at the right end of this product.

Now let's prove the previous proposition.
Case (i) R is a Euclidean domain:
We have already given the proof in this case in section §3, by diagonalizing the matrix using "elementary matrices" corresponding

to elementary row/column operations, and interpreting those as isomorphisms of R^m, R^n.

Case (ii) R is a p.i.d.
We will again prove the matrix can be diagonalized, but we will be obliged to introduce an additional class of matrices, which one may call "secondary matrices". The point is that although we do not have the Euclidean algorithm, and thus cannot construct a gcd by repeated subtraction using elementary operations, we still can find an invertible matrix which, given two non zero entries in a row or column, will replace one of them by the gcd of the two entries. The same diagonalization procedure as before will then work again.

Diagonalization algorithm: If there are no $\neq 0$ entries in the matrix $[\varphi]$, stop. Then $M \cong R^m$ is free, isomorphic to a product of copies of the free cyclic module R. If there is a $\neq 0$ entry, by interchanging rows and columns, bring one with the fewest number of prime factors to position (1,1), (upper left corner), and call it a. If there is another $\neq 0$ entry in the first column, bring it to position (2,1) by an interchange of rows, and call it b. Now we want to replace a by $c = \gcd(a,b)$ To do this first recall that since R is a p.i.d., the gcd of a,b is the generator of the ideal $(a,b) = (c)$. Hence we have $c = ax+by$ for some x,y, and $a = ca_1$, $b = cb_1$, for some a_1, b_1. Hence $1 = a_1x+b_1y$, and $(ab)/c = a_1 \cdot b = a \cdot b_1$. Consider the following matrix $\begin{bmatrix} x & y \\ -b_1 & a_1 \end{bmatrix}$. Note that this matrix multiplies the column vector $\begin{bmatrix} a \\ b \end{bmatrix}$ into the column vector $\begin{bmatrix} c \\ 0 \end{bmatrix}$ Thus if we let $[\sigma]$ be the square matrix with this 2×2 matrix in the upper left corner, and otherwise looking like the identity matrix, left multiplication of our matrix by $[\sigma]$ will replace the first two entries in the first column of $[\sigma]$ by the entries c, 0 [Since the 2×2 matrix above has $\begin{bmatrix} a_1 & -y \\ b_1 & x \end{bmatrix}$ as inverse, an inverse for $[\sigma]$ can be constructed analogously by placing this matrix in the upper left corner of the identity matrix.]
If there is another $\neq 0$ entry in the first column, say d, interchange rows to bring it to position (2,1). Repeating the previous procedure, replaces entry (1,1) by the gcd of c and d, and again makes entry (2,1) equal to 0. Proceeding as before, we eventually replace all

entries in the first column, except entry (1,1), by 0.
Then we work on the first row in an analogous way, then return if necessary to the first column, alternating as before. This time the process must end because the (1,1) entry is continually being replaced by a proper factor of the previous (1,1) entry. Since an element of a p.i.d. has only a finite number of prime factors, this process cannot go on for more steps than the number of prime factors of the original (1,1) entry. Thus eventually we have all zeroes in the first row and first column, except for entry (1,1).
Then we move to the submatrix of rows and columns after the first, and repeat the procedure. The rest of the argument proceeds in the same way as for **Z**. QED for existence.
This completes the proof of the decomposition theorem for finitely generated modules over a p.i.d.

Using the theorem, we can strengthen the lemma in which we proved that a submodule of a free finite module over a p.i.d. is free:
Corollary: If $N \subset R^n$ is a submodule of a finite free module over a p.i.d., then there is a free basis $\{x_1,....,x_n\}$ for R^n and a sequence of non units of R, $(d_1,....,d_s)$, with $d_i|d_{i+1}$, unique up to unit multiples, such that $\{d_1x_1, d_2x_2,...., d_sx_s\}$ is a free basis of N.
proof: In the proof of the theorem we showed how to realize a submodule N as the image of a map $\varphi: R^s \to R^n$, and how to find isomorphisms $\tau: R^n \to R^n$ and $\sigma: R^s \to R^s$ such that if $\tilde{\varphi} = (\tau \circ \varphi \circ \sigma)$, then $Im(\tilde{\varphi}) = d_1R \times ....d_sR \times \{0\} \times ....\{0\} \subset R^n$. It follows that the isomorphism $\tau^{-1}: R^n \to R^n$ carries the standard basis $\{e_j\}_{j=1....n}$ of R^n to a basis $\{x_j\}_{j=1...n}$ of R^n such that the basis $\{d_je_j\}_{j=1...s}$ of the submodule $Im(\tilde{\varphi})$ is carried to the basis $\{d_jx_j\}_{j=1....s}$ for N. QED.

Remark: Since a field is a p.i.d., the theorem gives another reminder that all finitely generated k-modules are free when k is a field, i.e. a finitely generated k-vector space is $\cong k^n$ for some n. [Since k has no ideals, a cyclic k-module is $\cong$ k, so a finite product of cyclic k-modules is $\cong k^n$.]

Example of the uniqueness proof procedure:
The point of uniqueness is to show that if $A = \prod_{i=1...t} R/(d_i)$, where $d_i|d_{i+1}$, and no d_i is a unit, then the prime factors of the d_i,

including multiplicities, can be recovered just from the isomorphism class of A. The method is to consider, for each prime p dividing the generator of the ideal ann(A), the dimension of the R/(p) vector spaces $(p^rA)(p)$, of elements of order p in the submodules p^rA.

Suppose $A \cong \mathbf{Z}_2\times\mathbf{Z}_6\times\mathbf{Z}_{18}\times\mathbf{Z}_{90}\times\mathbf{Z}_{360}$. Then the ideal ann(A) = $(360)\subset\mathbf{Z}$, hence its generator, is determined by the isomorphism class of A. Thus we recover the "largest" of the d_j, 360, or equivalently the smallest of the ideals (d_j), from the isomorphism class of A. Now we can recover all the prime factors occurring in all the d_j just by factoring this one. I.e. Since 2,6,18,90, all divide 360, we get all their prime factors, but not the multiplicities, from the prime factors of 360 = (8)(5)(9) = $(2^3)(5)(3^2)$. Now for each of these primes we ask for the submodule of elements of that order in A. The elements of order 2 in A, i.e. for which $(2)\subset\mathbf{Z}$ is the annihilator ideal, are those elements of the product $\mathbf{Z}_2\times\mathbf{Z}_6\times\mathbf{Z}_{18}\times\mathbf{Z}_{90}\times\mathbf{Z}_{360}$, such that every entry has order 2. Since in each factor there are exactly two such elements, eg. the elements of order 2 in $\mathbf{Z}_{90}$ are [0] and [45], this submodule is $A(2) \cong \mathbf{Z}_2\times\mathbf{Z}_2\times\mathbf{Z}_2\times\mathbf{Z}_2\times\mathbf{Z}_2$. Now we do not want to count the elements as we did when working only over $\mathbf{Z}$, since these modules will not necessarily be finite over any p.i.d., but this is a finite dimensional vector space over $\mathbf{Z}_2$, hence the dimension is determined by the isomorphism class of A. The $\mathbf{Z}_2$ dimension of this space, namely 5, gives us the number of factors in the product $A \cong \mathbf{Z}_2\times\mathbf{Z}_6\times\mathbf{Z}_{18}\times\mathbf{Z}_{90}\times\mathbf{Z}_{360}$, which are divisible by 2. Looking at the submodules of elements of orders 3,and 5, we get $\mathbf{Z}_3\times\mathbf{Z}_3\times\mathbf{Z}_3\times\mathbf{Z}_3$, and $\mathbf{Z}_5\times\mathbf{Z}_5$ respectively. Thus there are four d_j divisible by 3 (the last four), and two d_j divisible by 5 (the last two). Since there are more d_j divisible by 2 than by any other prime, the number of such d_j, namely 5, is the total number of factors d_j. So we know now there are 5 d_j in all, all five are divisible by 2, the last four are divisible by 3, and the last two are divisible by 5.
Next we want the exact multiplicities for each prime. To find the number of d_j which are divisible by 4, we look first at the module $2A \cong (A/A(2)) \cong \mathbf{Z}_3\times\mathbf{Z}_9\times\mathbf{Z}_{45}\times\mathbf{Z}_{180}$, and then at the submodule of elements of order 2 in this module. I.e. we look at $(2A)(2) \cong (A/A(2))(2) \cong \mathbf{Z}_2$, since the only $\neq 0$ element of order 2 here is ([0],[0],[0],[90]). This tells us that 2^2 is a factor of exactly one of the

d_i. Hence we now know there are five d_i, and in the first four of them 2 occurs with multiplicity exactly one. To see which d_i are divisible by 2^3, we look at $2^2A \cong A/A(2^2) \cong \mathbf{Z}_3\times\mathbf{Z}_9\times\mathbf{Z}_{45}\times\mathbf{Z}_{90}$, and then at the elements of order 2 in this module, $(2^2A)(2) \cong (A/A(2^2))(2) \cong \mathbf{Z}_2$, since the only $\neq 0$ element of order 2 is ([0], [0], [0], [45]). Thus the fifth d_i is also divisible by 2^3, so no d_i is divisible by exactly 2^2. We repeat this procedure again, and get $2^3A \cong A/A(2^3) \cong \mathbf{Z}_3\times\mathbf{Z}_9\times\mathbf{Z}_{45}\times\mathbf{Z}_{45}$, which has no nontrivial elements of order 2. Hence no d_i is divisible by 2^4, so the fifth d_i is divisible by exactly 2^3.

Moving on to the factor 3, we consider $3A \cong \mathbf{Z}_2\times\mathbf{Z}_2\times\mathbf{Z}_6\times\mathbf{Z}_{30}\times\mathbf{Z}_{120}$, whose submodule of elements of order 3 is $(3A)(3) \cong \mathbf{Z}_3\times\mathbf{Z}_3\times\mathbf{Z}_3$. Thus only the last three d_i are divisible by 3^2, hence d_2 is divisible by exactly 3^1. Next we consider $3^2A \cong \mathbf{Z}_2\times\mathbf{Z}_2\times\mathbf{Z}_2\times\mathbf{Z}_{10}\times\mathbf{Z}_{40}$, and hence $(3^2A)(3) \cong \{0\}$. Thus no d_i is divisible by 3^3, and so d_2, d_3, d_4 are divisible by exactly 3^2.

As for the prime 5, we know d_4, d_5 are divisible by 5. We have $5A \cong \mathbf{Z}_2\times\mathbf{Z}_6\times\mathbf{Z}_{18}\times\mathbf{Z}_{18}\times\mathbf{Z}_{72}$, $(5A)(5) \cong \{0\}$ and no d_i is divisible by 5^2.

Thus the sequences of prime power factors occurring in the five d_i are $(2,2,2,2,2^3)$, $(3^0,3,3^2,3^2,3^2)$, $(5^0,5^0,5^0,5,5)$. Multiplying these sequences together recovers the d_i, up to units, i.e. $d_1 = 2\cdot1\cdot1 = 2$, $d_2 = 2\cdot3\cdot1 = 6$, $d_3 = 2\cdot3^2\cdot1 = 18$, $d_4 = 2\cdot3^2\cdot5 = 90$, $d_5 = 2^3\cdot3^25 = 360$.

A memorable way to display this data is in a table of exponents for the various prime factors of each d_i. One can view the proof of uniqueness as the reconstruction of the exponents in this table from the dimensions of various p-torsion vector spaces associated to the given module.

	2	3	5
d_1	2^1	3^0	5^0
d_2	2^1	3^1	5^0
d_3	2^1	3^2	5^0
d_4	2^1	3^2	5^1
d_5	2^3	3^2	5^1

To recover the d_i from the table, just multiply together the entries in each row **End of example.**
In the next section we apply the cyclic decomposition theorem to finitely generated torsion modules over the ring R = k[X].

§7) The "Rational Canonical Form" of a matrix for a linear transformation of a vector space.

The work we put into proving the general structure theorem for finitely generated modules over p.i.d.'s will be amply justified in this section when we obtain a beautiful and useful structure theorem for the matrix of any linear transformation of a finite dimensional vector space over a field.
Let $f = a_0+a_1X+a_2X^2+\dots a_{r-1}X^{r-1}+X^r$ be any monic, non constant polynomial over a field k, and consider the following matrix C_f

associated to f: $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & 0 & 0 & -a_2 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 1 & 0 & -a_{r-2} \\ 0 & 0 & 0 & 0 & 1 & -a_{r-1} \end{bmatrix} = C_f$. This is called the

"companion matrix" of f.
We will prove that each matrix in $Mat_{n\times n}(k)$ is conjugate to exactly one "block" matrix in which the blocks along the diagonal are all companion matrices such that each associated polynomial divides

the next. i.e. one of form $\begin{bmatrix} [C_{f_1}] & 0 & 0 & 0 \\ 0 & [C_{f_2}] & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & [C_{f_n}] \end{bmatrix}$, where $f_1|f_2|\dots|f_n$.

First we review how a choice of basis yields a matrix for a map.

Bases, coordinate isomorphisms, and matrices

Recall that over any ring R, an $m\times m$ matrix is equivalent to an endomorphism $R^m \to R^m$, and that an R module M is free of rank m,

iff M has a free basis of m elements $(x_1,\ldots,x_m)$. We want to recall how an endomorphism of any free module M can be represented by a matrix, once an (ordered) basis is chosen. We may be careless, as is often the case, and say "basis" when we mean "ordered basis".

Lemma: If M is a free, rank m, R-module, then an *ordered* basis of M is equivalent to an isomorphism $\sigma:R^m \to M$.
proof: Since (e_j) is a basis of R^m, and an isomorphism carries a basis to a basis, any isomorphism $\sigma:R^n \to M$ carries the ordered basis (e_j) to an ordered basis (x_j) of M, where $x_j = \varphi(e_j)$. Conversely, if $(x_j)_{j=1,\ldots,m}$ is an ordered basis of M, the map $\sigma:R^m \to M$ defined by $\varphi(\alpha_1,\ldots,\alpha_m) = \Sigma\alpha_j x_j$, is a homomorphism by the usual properties of addition and multiplication, is injective since $(x_j)_{j=1,\ldots,m}$ is independent, and surjective since $(x_j)_{j=1,\ldots,m}$ generates M. QED.

Remark: The isomorphism $\sigma:R^m \to M$ is sometimes called a parametrization of M, and the inverse isomorphism $\sigma^{-1}:M \to R^n$ a coordinate map for M. In particular, if x is an element of M, the vector $\sigma^{-1}(x)$ in R^n is called the coordinate vector of x, with respect to the basis (x_j) of M giving rise to the isomorphism.

The matrix of an endomorphism, associated to a basis
If $\varphi:M \to M$ is any endomorphism of a free rank m module M, and $S = (x_j)_{j=1,\ldots,m}$ is an ordered basis of M, by the previous lemma we have mutually inverse isomorphisms $\sigma:R^m \to M$, and $\sigma^{-1}:M \to R^m$. Hence the composition $\tilde{\varphi} = (\sigma^{-1}\circ\varphi\circ\sigma):R^m \to R^m$ has a matrix $[\tilde{\varphi}]$, which we may also write as $[\varphi]_S$ or simply $[\varphi]$, if the basis used to construct it seems unimportant. This is the matrix of φ determined by the ordered basis S. You should convince yourself that the jth column of $[\varphi]$ is the coordinate vector of $\varphi(x_j)$ with respect to the basis (x_j). I.e. $[\varphi] = [a_{ij}]$ iff for all j, $\varphi(x_j) = a_{1j}x_1 + \ldots + a_{nj}x_m$.
Now it frequently happens that $M=R^m$ so that φ already has a "standard" matrix, namely the one determined by the standard basis (e_j); but this matrix may not be particularly useful, so we may wish to choose another basis which yields a nicer matrix for φ. [A matrix is nicer for example if it is easier to calculate with, or if it reveals more easily the properties of the map φ it represents.] In this case, the isomorphisms σ, σ^{-1} will themselves be given by

invertible matrices, say A, A^{-1}, and we see that the new matrix for φ, $[\tilde{\varphi}] = A^{-1}[\varphi]A$, the one associated to the new basis $\{x_j\}$, is conjugate to the original one. [You should check that the jth column of A is the standard coefficient vector for the jth vector x_j of the new basis of R^m.] In this case, the set of matrices for φ obtained from all possible bases of R^m, is precisely the conjugacy class of $[\varphi]$ under the action of $GL_m(R) \cong Aut_R(R^m)$ on the ring $Mat_{m\times m}(R) \cong End_R(R^m)$.

Exercise #140) Let T be the linear transformation $T:\mathbb{Q}^2 \to \mathbb{Q}^2$ whose matrix in the standard basis is $\begin{bmatrix} 3 & 5 \\ 2 & -4 \end{bmatrix}$. Show all calculations and find the matrix for T w.r.t. the ordered basis {(1,-2), (3,-1)}.

Remark on terminology: The relation $B \approx A^{-1}BA$, which we have called "conjugacy", is usually called "similarity" in linear algebra. One difference which may justify the distinction is that in linear algebra we do not require B to be invertible, just A, in this relation. I.e. we have a proper inclusion $Aut_R(R^m) \subset End_R(R^m)$, of the group of units into a ring, and the units are acting by conjugation on the whole ring, not just on the group elements. So more accurately, we should say above that two matrices represent the same linear transformation, but in different bases, iff the matrices are similar.

The application we present next of these ideas will exploit the relationship between the k vector space structure and the k[X] module structure of a finite torsion k[X] module.

k[X]-modules, viewed as pairs (M,T)

Let M be a finitely generated k[X] module, i.e. let M be an abelian group plus a ring homomorphism $k[X] \to End(M)$, where k is a field. Then the restriction of this ring map to k gives a ring map $k \to End(M)$, hence a k-vector space structure on M. More simply, if you can multiply elements of M by any elements of k[X], then in particular you can multiply by elements of k. The subring $End_k(M) \subset End(M)$ is the ring of k-vector space endomorphisms of M, i.e. the group endomorphisms that commute with multiplication by elements of k. Since k[X] is a commutative ring and $k[X] \to End(M)$ is

a ring homomorphism, X commutes with elements of k both in k[X] and in End(M), i.e. k[X] actually maps into the subring $\mathrm{End}_k(M)$. Thus not only is M a k-vector space, but the image of X in $\mathrm{End}_k(M)$ is a distinguished k-linear transformation T of M. Since conversely, if M is a k-vector space and $T:M\to M$ a k-linear endomorphism of M, there is a unique ring map $k[X]\to \mathrm{End}_k(M)$ taking c in k to c·Id, and X to T, a k[X] module is equivalent to a pair (M,T) where M is a k-vector space, and T is a k-endomorphism of M.

If M is a k[X] module and $N\subset M$ is a submodule, then N is closed not only under multiplication by elements of k, but also under multiplication by X, i.e. under the action of T. In other words, $T(N)\subset N$, and N is a "T-invariant subspace" of M. Thus if a k[X] module is interpreted as a pair (M,T) as above, then a k[X] submodule of M is a T-invariant subspace $N\subset M$. Finally, a cyclic k[X] module is a pair (M,T) where $M \cong k[X]/(f)$ for some polynomial f. This is a k[X] isomorphism, which means it is a k-vector space isomorphism such that multiplication by X in k[X]/(f) corresponds to action by T on M. I.e. if v in M corresponds to [g] in k[X]/(f), then T(v) corresponds to [X·g] in k[X]/(f). In this case, if v in M corresponds to [1] in k[X]/(f), then $T^r(v)$ corresponds to $[X^r]$. Thus if f has degree s, the standard k-basis $\{[1],[X],\ldots,[X^{s-1}]\}$ of k[X]/(f), corresponds to the k-basis $(v, Tv, T^2v,\ldots,T^{s-1}v)$ of M. A cyclic k[X] module M is a k-vector space M with an endomorphism T which admits a basis of this form. [If M has such a basis, and if $T^sv = -\alpha_0 v - \alpha_1 Tv - \ldots - \alpha_{s-1}T^{s-1}v$, then $M \cong k[X]/(f)$, as k[X] module, where $f = \alpha_0+\alpha_1X+\ldots+\alpha_{s-1}X^{s-1}+X^s$.] If f=0, then $M\cong k[X]$, and has an infinite k-basis of form $(v,Tv,T^2v,\ldots,T^sv,\ldots)$.

The minimal polynomial of a finite dim'l k-endomorphism: Now let M be a finitely generated k[X] module, which is thus isomorphic as k[X] module, and thus also as k-vector space, to some finite product of cyclic k[X] modules $M \cong (k[X])^m \times \prod_j k[X]/(f_j)$. Since the factors of form k[X] are infinite dimensional over k, and those of form $k[X]/(f_j)$ are finite dimensional, this is a finite dimensional k-vector space if and only if as a k[X] module, it is a torsion module. Thus the study of pairs (M,T) where M is a finite dimensional k-vector space, and T is an endomorphism of M, is equivalent to the study of finite torsion k[X] modules.

If $M \cong \Pi_{j=1,\dots,n}\, k[X]/(f_j)$ is a torsion $k[X]$ module, where each $f_j | f_{j+1}$, then we know the annihilator ideal of M is $(f_n) \subset k[X]$. Since multiplication by X corresponds to the endomorphism $T: M \to M$, and multiplication by $f_n(X)$ kills every element of M, this means that the k-endomorphism $f_n(T): M \to M$ is identically zero. We say that T satisfies the polynomial $f_n(X)$, and since this is the polynomial of least degree that T satisfies, f_n is called the "minimal polynomial" of T. Thus given a pair (M,T), where M is finite dimensional as a k-vector space, the minimal polynomial of T is the monic generator of the kernel of the associated k-algebra map $k[X] \to \mathrm{End}_k(M)$ which takes $X \mapsto T$, and from the decomposition above we know it has degree $\leq$ the k-dimension of M.
Recall that the annihilator of a finite abelian group M is the least common multiple of the orders of all the elements, and if $M \cong \Pi_{i=1,\dots,s}\, \mathbb{Z}/n_i$, then $\mathrm{ann}(M) = n_s$. In our case, every $\neq 0$ element v of the fin. gen. torsion k[X]-module (M,T) is killed by multiplication by some monic polynomial f, of least degree in k[X]. If such an f is called the order of v, then the minimal polynomial for T is the monic least common multiple of the orders of all v in M. If $M \cong k[X]/(f)$ is a cyclic torsion k[X] module with generator v, then the order of $v = f$ = the annihilator of M = the minimal polynomial of T.

The rational canonical matrix of an endomorphism
Let (M,T) be any finite dimensional k-vector space, plus an endomorphism T, viewed as a k[X] module, and let $M \cong \Pi_j\, k[X]/(f_j)$ be an isomorphism as k[X] modules, where each $f_j | f_{j+1}$. Then each submodule $k[X]/(f_j) \cong N_j$ corresponds to a k[X] submodule of M, i.e. a T invariant subspace $N_j \subset M$. To get the "rational canonical" form of the matrix for T, we will choose in each subspace N_j, the k-basis corresponding to the standard k-basis of $k[X]/(f_j)$.
So let $f_1 = a_0 + a_1X + \dots + a_{r-1}X^{r-1} + X^r$, and let $\{v, Tv, T^2v, \dots, T^{r-1}v\}$ be the k-basis of N_1 corresponding to the k-basis $[1], [X], \dots [X^{r-1}]$ of $k[X]/(f_1)$. To form the matrix of T_j = the restriction of T to N_j, associated to this matrix, we let T act on the jth basis vector, and expand the result in terms of this basis, and put the coefficients in the jth column of the matrix. Thus the first basis vector is v, apply T to get Tv, and expand as $Tv = 0 \cdot v + 1 \cdot Tv + 0 \cdot T^2v + \dots + 0 \cdot T^{r-1}v$, so the coefficient vector is (0,1,0,....,0), and this is the first column of

the matrix The second basis vector is Tv, applying T gives T^2v, which has expansion $T^2v = 0\cdot v + 0\cdot Tv + 1\cdot T^2v + 0\cdot T^3v + \ldots + 0\cdot T^{r-1}v$, so the second column vector is (0,0,1,0,....,0). Continuing in this way, the r-1st column vector is (0,....,0,0,1). The last column looks different from the others, since the last basis vector vector is $T^{r-1}v$, applying T gives T^rv, and since we know $f_1(T)$ annihilates N_1, the endomorphism $a_0I + a_1T + a_2T^2 + \ldots + a_{r-1}T^{r-1} + T^r$ is identically zero on N_1. In particular $a_0v + a_1Tv + a_2T^2v + \ldots + a_{r-1}T^{r-1}v + T^rv = 0$, so that $T^rv = -a_0v - a_1Tv - a_2T^2v - \ldots - a_{r-1}T^{r-1}v$. Thus the last column is the coordinate vector $(-a_0, -a_1, -a_2, -\ldots, -a_{r-1})$. Therefore the matrix of T_j in this basis is precisely the matrix we called the "companion matrix" of f_1:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & 0 & 0 & -a_2 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 1 & 0 & -a_{r-2} \\ 0 & 0 & 0 & 0 & 1 & -a_{r-1} \end{bmatrix} = [C_{f_1}]$$

Continuing, we do this for each j = 1,....n, and take as our basis for M, the union of these bases for the N_j. Then the associated "rational canonical" matrix for T is the following block matrix:

$$\begin{bmatrix} [C_{f_1}] & 0 & 0 & 0 \\ 0 & [C_{f_2}] & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & [C_{f_n}] \end{bmatrix}, \text{ where } f_1|f_2|\ldots|f_n.$$

It follows from what we have said above, that every m×m matrix T is similar (i.e. conjugate) to an m×m matrix in this form. Conversely, if T is similar to such a matrix, then the pair (k^m,T) is isomorphic as k[X] module to the product $\prod_{j=1\ldots n} k[X]/(f_j)$. By the uniqueness for the decomposition theorem, a given (m×m) matrix T is conjugate to exactly one (m×m) matrix in rational canonical form, i.e. to exactly one block matrix, whose blocks are companion matrices, such that the associated polynomials divide each other. In particular we have a unique explicit representative for each conjugacy class of the group $GL_m(k) = Aut_k(k^m)$.

Exercise #141) Write down the 3×3 companion matrix C of the polynomial $f(X) = X^3$, and prove by direct computation that X^3 is the minimal polynomial of C. I.e. prove that $C^3 = [0]$, and that unless α,β,γ, are all zero in k, then $\alpha\cdot Id + \beta\cdot C + \gamma\cdot C^2 \neq [0]$.

Example of rational canonical form:
Suppose we can show that the pair (M,T) is isomorphic as k[X] - module to the following product of cyclic k[X] - modules:
$(k[X]/(X-2)) \times (k[X]/((X-2)(X-3)^2) \times k[X]/((X-2)^2(X-3)^2)$.
Then there are three "invariant factors" for (M,T), which determine (M,T) completely: $f_1 = X-2$, $f_2 = (X-2)(X-3)^2$, and $f_3 = (X-2)^2(X-3)^2$. The last of these, $f_3 = (X-2)^2(X-3)^2$, is the minimal polynomial for T, which we may denote μ_T, or $\mu_T(X)$. To get the rational canonical form of [T], we first multiply out the invariant factors to find their coefficients: $f_1 = X-2$, $f_2 = X^3-8X^2+21X-18$, and $f_3 = X^4-10X^3+37X^2-60X+36$, if I haven't made a mistake. Then the rational canonical form is a block matrix with three blocks, each block being the companion matrix for one of these polynomials. Remembering that the right-most column is the *negative* of the coefficients of the polynomials, we get the following 8×8 matrix for [T], where all blank spaces are filled in by zeroes:

$$\left[\begin{array}{cccccccc} 2 & & & & & & & \\ & 0 & 0 & 18 & & & & \\ & 1 & 0 & -21 & & & & \\ & 0 & 1 & 8 & & & & \\ & & & & 0 & 0 & 0 & -36 \\ & & & & 1 & 0 & 0 & 60 \\ & & & & 0 & 1 & 0 & -37 \\ & & & & 0 & 0 & 1 & 10 \end{array}\right] = [T]$$

In the next section we will see how a slight variation on this theme yields the simpler, more informative, "Jorden" matrix for T.