Similarity of Matrices via Module Theory

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

Abstract

The aim of this article is to give an introduction to the structure theorem of finitely generated modules over a PID confining ourselves only to $F[X]$-modules. The treatment acquaints the reader with all basic results such as invariant factor theorem, elementary divisor theorem, rational and Jordan canonical forms. The entire theory (up to Theorem 31) can be carried out almost verbatim for finitely generated modules over a Euclidean domain. In particular, the analogues of Theorem 25 and Theorem 31 together yield the structure theorem for finitely generated abelian groups. In fact, we encourage the reader to interpret, formulate the analogous results in this case and prove them as he goes through the article. I hope that this will motivate the readers to plunge into a more leisurely and detailed study of modules.

Definition 1. Let R be a commutative ring with identity. Let M be an abelian group (the group operation being written additively). Assume that there is a map $R \times M \to M$ given by $(a, x) \mapsto ax$ with the following properties:

1. $(a + b)x = ax + bx$,
2. $a(x + y) = ax + ay$,
3. $ab(x) = a(bx)$,
4. $1x = x$,

for all $a, b \in R$ and $x, y \in M$. Then M is called an R-module or a module over R.

The typical examples are given below.

Example 2. If R is a field, then R-modules are nothing other than vector spaces over R.

Example 3. Let M be an abelian group written additively. Let $R = \mathbb{Z}$ be the ring of integers. Then M is an R-module via the map

$$(n, x) \mapsto nx = \begin{cases} x + \cdots + x, n\text{-times, if } n > 0 \\ 0, \text{if } n = 0 \\ (-x) + \cdots + (-x), -n\text{-times, if } n < 0. \end{cases}$$

Example 4. If S is a subring of R, then R is an S-module in an obvious way.

Example 5. If I is an ideal of R, then I is an R-module in an obvious way.

Example 6. Let $M := R^n$ denote the n-fold product of R. Then if we define $a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n)$, then M is an R-module.

Example 7. This is the most important example for our purpose. Let V be a vector space over a field F and $f\colon V \to V$ be a linear map. Let $F[X]$ denote the polynomial ring over F. Given a polynomial $p(X) := a_0 + a_1X + \cdots + a_dX^d$, we make V a module over $F[X]$ by setting

$$p(X)v \equiv pv := (a_0I + a_1f + \cdots a_df^d)v.$$

For example, if $p(X) = 1 + 2X - 4X^3$, then $pv = v + 2f(v) - f^3(v)$. One easily verifies that with this operation (of course, along with the vector addition!) V becomes an $F[X]$-module.

To show the dependence on f of the module structure on V, we denote this module by V_f.

Remark 8. In an R-module M, it is possible that $ax = 0$ with neither x being the additive identity 0 in M nor a being the zero element of R. Give examples of such a phenomenon in the case of Examples 3 and 7.

Definition 9. Let $S \subset M$ where M is an R-module. The set $\operatorname{Ann}(S) := \{a \in R : ax = 0 \text{ for all } x \in S\}$ is called the *annihilator* of S.

If $S = \{v\}$, then we denote $\operatorname{Ann}(S)$ by $\operatorname{Ann}(v)$.

If $\operatorname{Ann}(v) \neq (0)$, then we say that v is a *torsion element.* If $0 \in M$ is the only element which is torsion, then M is said to be *torsion-free.*

Ex. 10. Show that $\operatorname{Ann}(S)$ is an ideal in R.

Ex. 11. (a) What are the torsion elements in V_f? When is V_f torsion free? (b) What are the torsion elements in a $\mathbb{Z}$-module?

Ex. 12. Think of interesting subsets S in Examples 3 and 7 and find their annihilators.

Ex. 13. If $m(X)$ is the minimal polynomial of $f\colon V \to V$, then $\operatorname{Ann}(V_f)$ is the principal ideal $(m(X))$ in $F[X]$.

Ex. 14. (i) Define a submodule of an R-module M.
Show that $W \subset V$ is an $F[X]$-submodule of V_f iff W is an f-invariant subspace.
What are the submodules of R considered as a module over itself?
What are the submodules of an abelian group considered as a module over $\mathbb{Z}$?
(ii) Define the quotient module of M with respect to a submodule N of M.
(iii) Define an R-module homomorphisms between two R-modules M_i, $i = 1, 2$.
(iv) What do you mean by saying that a subset S of M is a *set of generators* for M over R?
(v) When do you say an R-module is finitely generated?
(vi) Is V_f finitely generated?

Definition 15. An R-module M is said to be *cyclic* if it is generated by a single element.

Ex. 16. Give examples of cyclic R-modules in the case of Examples 3 and 7.

Proposition 17. *Let $M = [v]$ be a cyclic $F[X]$-module whose annihilator* $\operatorname{Ann}(M)$ *is the ideal generated by $p(X)$ of degree d. Then M has a natural structure of a vector space over F of dimension d.*

Proof. The reader is urged to prove this on his own.

The elements $v, Xv, \cdots, X^{d-1}v$ are linearly independent over F, by the very definition of the polynomial p. Thus the F-span $\{v, Xv, \ldots, X^{d-1}v\} = M$. Since M is an $F[X]$-module, it is a module over F also and it has $\{v, Xv, \ldots, X^{d-1}v\}$ as a basis. Hence the result. □

Question 18. What is the analogue of the last result in the case of an abelian group considered as a $\mathbb{Z}$-module?

Ex. 19. Formulate and prove First isomorphism theorem for module homomorphisms.

Proposition 20. *If M is an cyclic R-module, then M is isomorphic to $R/\mathrm{Ann}\,(M)$ as R-modules.*

Proof. The reader should prove this on his own.

Let $M = [v]$. Consider the map $f\colon R \to M$ given by $f(r) = rv$. Then f is onto and the kernel of f is $\mathrm{Ann}\,(v) = \mathrm{Ann}\,(M)$. The result follows from the first isomorphism theorem. □

Question 21. What is the analogue of the last result in the case of an abelian group considered as a $\mathbb{Z}$-module?

Ex. 22. Deduce Proposition 17 from the last proposition.

Ex. 23. Show that any torsion free cyclic R-module is isomorphic to R itself. What is the analogue of this in the case of an abelian group considered as a $\mathbb{Z}$-module?

Recall that two linear maps $f, g\colon V \to V$ are *similar* iff there exists an automorphism $\varphi\colon V \to V$ such that $g = \varphi^{-1} \circ f \circ \varphi$.

Ex. 24. Let $f, g\colon V \to V$ be linear maps of V. Then show that f and g are similar iff the modules V_f and V_g are isomorphic as $F[X]$-modules.

Theorem 25. *Let M be a finitely generated $F[X]$-module. Then M can be decomposed into a direct sum of cyclic submodules*

$$M = [v_1] \oplus \cdots \oplus [v_k], \text{ where } \mathrm{Ann}\,(v_1) \subseteq \mathrm{Ann}\,(v_2) \subseteq \cdots \subseteq \mathrm{Ann}\,(v_k) \text{ and } \mathrm{Ann}\,(v_k) \neq F[X].$$

Furthermore, the ideals $\mathrm{Ann}\,(v_i)$ are uniquely determined by M.

Proof. We start with an easy observation: If $w_1, \ldots, w_l$ generate M, so do $w_1, \ldots, w_{l-1}, w_l + qw_j$ for $q \in F[X]$ and $1 \leq j \leq l-1$.

Let k be the smallest number of elements required to generate M.

Easy Case: Suppose that we can find a set $\{v_1, \ldots, v_k\}$ of generators among which no non-trivial relation holds, that is, if $a_1v_1 + \cdots + a_kv_k = 0$ implies that $a_i = 0$ for $1 \leq i \leq k$. Then, we have,

$$M = [v_1] \oplus \cdots \oplus [v_k]$$

and $\mathrm{Ann}\,(v_i) = (0)$ for each i. The first part of the theorem is now established in this case.

The left out case is when *any* set of k generators gives rise to a non-trivial relation among them. Then there exists a system $\{v_1, \dots, v_k\}$ of generators for which a relation of the form $a_1v_1 + \cdots + a_kv_k$ holds with a nonzero coefficient a_i of minimum degree. (Go through this assumption once again to understand it properly! We shall refer to this as the minimality assumption.) We may assume that $i = k$.

Claim 1. Each a_j is a multiple of a_k for $1 \leq j \leq k-1$, say, $a_j = q_ja_k$.

Reason: For, if $a_j = q_ja_k + r_j$ with $\deg r_j < \deg a_k$, then

$$a_1v_1 + \cdots + r_jv_j + \cdots + a_k(v_k + q_jv_j) = 0.$$

Thus, we see that $\{v_1, \cdots, v_j \cdots, v_k + q_jv_j\}$ is a set of generators with a non-trivial relation. If $r_j \neq 0$, then this would contradict our minimality assumption on a_k.

Claim 2. We claim that if $b_1v_1 + \cdots + b_kv_k = 0$, then b_k is a multiple of a_k.

Reason: For, if $b_k = qa_k + r$ with $\deg r < \deg a_k$, then by subtracting q times the first relation from the second we obtain

$$(b_1 - qa_1)v_1 + \cdots + rv_k = 0.$$

As in the last claim, we deduce that $r = 0$.

Using the notation of Claim 1, if we set $w_k := q_1v_1 + \cdots q_{k-1}v_{k-1} + v_k$, then Claim 1 shows that $a_kw_k = 0$.

Let $M' := [v_1, \dots, v_{k-1}]$, then $M = M' \oplus [w_k]$. (Why?)

Reason: Since the relation $b_1v_1 + \cdots + b_{k-1}v_{k-1} + b_kw_k = 0$ implies that b_k is a multiple of a_k by Claim 2. Since $a_kw_k = 0$, we see that $b_kw_k = 0$. Hence both the summands $b_1v_1 + \cdots + b_{k-1}v_{k-1}$ and b_kw_k are zero. Hence the sum is direct.

We are now ready to prove the first part of the theorem in the second case by induction on k. We can write M' as a direct sum

$$M' = [w_1] \oplus \cdots \oplus [w_{k-1}],$$

with $\operatorname{Ann}(w_1) \subset \cdots \subset \operatorname{Ann}(w_{k-1})$.

It remains to prove that $\operatorname{Ann}(w_{k-1}) \subset \operatorname{Ann}(w_k)$. If $q \in \operatorname{Ann}(w_{k-1})$, then $qw_{k-1} = 0$ so that $qw_{k-1} + a_kw_k = 0$. But then by Claim 1, it follows that a_k divides q and hence $q \in \operatorname{Ann}(w_k)$. This completes the proof of the first part of the theorem.

We now attend to uniqueness. Let us assume that r of the ideals $\operatorname{Ann}(v_i)$ are zero $(0 \leq i \leq r)$. Define

$$\begin{aligned} M_1 &:= [v_1] \oplus \cdots \oplus [v_r] \\ M_2 &:= [v_{r+1}] \oplus \cdots \oplus [v_k]. \end{aligned}$$

Clearly M_2 is the set of *torsion elements*, that is,

$$M_2 = \{x \in M : \operatorname{Ann}(x) \neq (0)\}.$$

> For, if $x = p_1v_1 + \cdots + p_kv_k \in M$ is such that $qx = 0$ for some nonzero $q \in F[X]$, since the sum is direct, it follows that $qp_1v_1 = \cdots = qp_kv_k = 0$. Since $\mathrm{Ann}\,(v_i) = 0$ for $1 \leq i \leq r$, we deduce that $qp_1 = \cdots = qp_r = 0$. Therefore, $p_i = 0$, for $1 \leq i \leq r$.

Thus, M_2 is uniquely determined by M.

Using the decomposition $M = M_1 \oplus M_2$, we can write any $x \in M$ in the form $x = x_1 + x_2$ with $x_i \in M_i$, $i = 1, 2$. The map $x \mapsto x_1$ is an R-module homomorphism of M onto M_1 with kernel M_2. So, M_1 is isomorphic to M/M_2, by the first isomorphism theorem (Ex. 19).[1]

Thus the proof of the uniqueness of the invariant factors is now reduced to the special cases where M consists of torsion elements alone or where M is *torsion-free*. The first case will be attended to after Theorem 31.

Let us therefore assume that $M = [v_1] \oplus \cdots [v_k]$ where $\mathrm{Ann}\,(v_i) = (0)$ for $1 \leq i \leq k$. We have to show that k is uniquely determined by M. The idea is to show that k is the maximum number of *linearly independent elements* over $F[X]$. (What is the meaning of the italicized phrase?)

> Reason: We say that distinct elements $w_1, \ldots, w_r \in M$ are linearly independent over $F[X]$ if every relation of the form $f_1w_1 + \cdots + f_rw_r = 0$ implies that $f_i = 0$ for $1 \leq i \leq r$. The map
>
> $$p_1w_1 + \cdots + p_rw_r \mapsto (p_1, \ldots, p_r)$$
>
> is an $F[X]$-isomorphism of M onto $F[X]^r$. This maps therefore preserves linear independence over $F[X]$. Let $F(X)$ denote the field of quotients f/g where $f, g \in F[X]$ with $g \neq 0$. Then the linear independence of elements of $F[X]^r$ over $F[X]$ is equivalent to the their linear independence as elements of the vector space $F(X)^r$ over the field $F(X)$. (Why? Because, the denominators can be cleared!)

Thus k is the maximum number of *linearly independent elements* of M over $F[X]$. Therefore, it is an invariant of M. □

Ex. 26. Formulate the analogue of the last result in the case of an abelian group considered as a $\mathbb{Z}$-module. Modify the proof above to prove your formulation.

Definition 27. The ideals $\mathrm{Ann}\,(v_i)$ or their monic generators are called *invariant factors* of M.

Ex. 28. Let $M = \mathbb{Z} \oplus \mathbb{Z}_2$. Let $x = (1, 0)$, $y = (0, 1)$, $x' = (1, 1)$. Then $M = [x] \oplus [y]$ as well as $M = [x'] \oplus [y]$.

Proposition 29. *Let M be a cyclic $F[X]$-module. Let the annihilator* $\mathrm{Ann}\,(M)$ *be generated by a monic polynomial $m(X)$. Assume that p_i, $1 \leq i \leq r$, be the distinct monic irreducible factors of m in $F[X]$. Then*

$$M = [v_1] \oplus \cdots \oplus [v_r],$$

where $\mathrm{Ann}\,(v_i)$ *is the ideal generated by a power of p_i, $1 \leq i \leq r$.*

[1] Note that we do *not* claim that M_1 is uniquely determined by M. See Ex. 28.

Proof. Let $M = [v]$. Assume that $m = fg$ where f and g are relatively prime. If $w := fv$, then $\mathrm{Ann}\,(w) = (g)$: for $qw = 0$ implies $qf \in (m)$ so that m divides qf, that is, g divides q. Similarly, if $u = gv$, then $\mathrm{Ann}\,(u) = (f)$.

Since f and g are relatively prime, there exist $a, b \in F[X]$ such that $af + bg = 1$. This implies that $[v] = [w] + [u]$. (Check!) Also, any element $x \in [w] \cap [u]$ is annihilated by both by f and g. It follows that $[w] \cap [u] = (0)$ and so the sum $[v] = [w] + [u]$ is direct.

The general case follows by induction on the number of distinct irreducible factors of p. □

Ex. 30. Formulate the analogue of the last result in the case of an abelian group considered as a $\mathbb{Z}$-module. Modify the proof above to prove your formulation.

Theorem 31. *Let M be a finitely generated $F[X]$-module. Assume that its annihilator $\mathrm{Ann}\,(M)$ is generated by a nonzero monic polynomial $m(X)$. If p_i, $1 \le i \le r$, are the distinct monic irreducible factors of $m(X)$, then*

$$M = K_1 \oplus \cdots \oplus K_r,$$

where each K_i is the submodule of all elements in M annihilated by some power of p_i.

Moreover, each K_i may be expressed as a direct sum

$$K_i = [v_{i1}] \oplus \cdots \oplus [v_{ik_i}],$$

where $\mathrm{Ann}\,(v_{i1}) \subset \mathrm{Ann}\,(v_{i2}) \subset \cdots \subset \mathrm{Ann}\,(v_{ik_i})$.

The ideals $\mathrm{Ann}\,(v_{ij})$, *(*$1 \le i \le r$, $1 \le j \le k_i$*), are uniquely determined by M.*

Proof. The proof is a typical case of book-keeping exercise using Theorem 25 and Proposition 29.

We decompose M into a direct sum of cyclic submodules each of which is annihilated by some power of an irreducible factor of m. The set K_i of all "vectors" annihilated by some power of p_i is therefore the direct sum of those cyclic submodules that are annihilated by some power of p_i. We may enumerate these submodules as

$$[v_{i1}], \ldots, [v_{ik_i}] \text{ where } \mathrm{Ann}\,(v_{ij}) = [p_i^{n_{ij}}] \text{ and } n_{i1} \ge n_{i2} \ge \cdots \ge n_{ik_i}.$$

Since the submodules K_i are uniquely determined by M, to establish the uniqueness of elementary divisors, we can restrict ourselves to the case when m is a power of an irreducible polynomial, say p. Then

$$M = [v_1] \oplus \cdots \oplus [v_k] \text{ where } \mathrm{Ann}\,(v_j) = (p^{n_j}) \text{ and } n_1 \ge n_2 \ge \cdots \ge n_k.$$

In particular, $\mathrm{Ann}\,(M) = (p^{n_1})$.

We now prove the result by induction on n_1. The map $v \mapsto pv$ is an $F[X]$-homomorphism of M to M. Let K be the kernel of this homomorphism. Then $\sum_{i=1}^{k} q_i v_i \in K$ iff each q_i is a multiple of $p^{n_i - 1}$, $1 \le i \le k$. Hence K is the direct sum of k cyclic submodules each of which has (p) as its annihilator. By Proposition 17, we see that

$$\dim_F K = k \times \deg p.$$

Therefore, k is uniquely determined by M. Now p^{n_1-1} generates $\operatorname{Ann}(pM)$ and so by induction hypothesis, the integers

$$n_1 - 1 \geq n_2 - 1 \geq \cdots \geq n_l - 1 > 0$$

are determined uniquely. Finally the sequence

$$n_1 \geq n_2 \geq \cdots \geq n_l > n_{l+1} = \cdots = n_k = 1$$

is uniquely determined by M. □

Completion of the proof of Theorem 25. The last part of Theorem 25 now follows since the invariant factor $\operatorname{Ann}(v_i)$ of the said theorem is generated by $\prod_i p_i^{n_{ij}}$ where the product is taken over all i for which n_{ij} is defined.

Ex. 32. Formulate the analogue of the last result in the case of an abelian group considered as a $\mathbb{Z}$-module. Modify the proof above to prove your formulation.

Definition 33. The ideals $\operatorname{Ann}(v_{ij})$ are called the *elementary divisors* of M.

In the case when $M = V_f$, the invariant factor and elementary divisors are called namely the invariant factors of f and elementary divisors of f.

Proposition 34. *Two finitely generated modules over $F[X]$ are isomorphic iff they have the same invariant factors.*

Two finitely generated torsion modules over $F[X]$ are isomorphic iff they have the same elementary divisors. □

Proof. This is an immediate consequence of Theorems 25 and 31 and Proposition 20. □

Question 35. What is the analogue of the last result in the case of abelian groups?

Lemma 36. *If V_f is a cyclic $F[X]$-module with generator, say, v, then V has an ordered basis with respect to which f has a matrix*

$$B = \begin{pmatrix} 0 & 0 & 0 & \dots & & a_1 \\ 1 & 0 & 0 & \dots & & a_2 \\ 0 & 1 & 0 & \dots & & a_3 \\ & & \ddots & \ddots & & \\ 0 & 0 & 0 & 1 & 0 & a_{d-1} \\ 0 & 0 & 0 & & 1 & a_d \end{pmatrix}$$

where $m_v(X) = X^d - a_1 X^{d-1} - \cdots - a_d$.

Proof. Proposition 17 says that $\{v, f(v), \ldots, f^{d-1}(v)\}$ is an ordered basis of V. Clearly, with respect to this basis, the matrix is as displayed above. □

The matrix B of the lemma is called the *companion matrix* of the polynomial $m_v(X)$.

Theorem 37. *There exists an ordered basis of V with respect to which f has a matrix* $\operatorname{diag}(B_1, \ldots, B_k)$ *where each B_j is of the form B of Lemma 36.*

Proof. This is an immediate consequence of Theorem 25. □

The matrix of the last theorem is called the *rational canonical form* of f.

If we assume that the minimal polynomial $m(X)$ of f can be factorized into linear factors in $F[X]$, then we can arrive at a simpler matrix representation of f.

Lemma 38. *If V_f is a cyclic module, say, $[v]_f$, where $m_v(X) = (X - \lambda)^n$, then there is an ordered basis of V with respect to which f is represented by the matrix $J_n(\lambda)$, where*

$$J_n(\lambda) := \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & & \\ & & \ddots & 1 & \\ & & & & \lambda \end{pmatrix}.$$

Proof. It follows from the definition of the minimal polynomial that the set of vectors

$$\{v, (X - \lambda)v, \ldots, (X - \lambda)^{d-1}v\}$$

is linearly independent. Let $v_i := (X - \lambda)^{i-1}v$. Then we have

$$(X - \lambda)v_i = v_{i+1}, \text{ for } 1 \le i \le n - 1 \text{ and } (X - \lambda)v_n = 0.$$

That is,

$$\begin{aligned} fv_1 &= \lambda v_1 + v_2 \\ &\vdots \\ f(v_{n-1}) &= \lambda v_{n-1} + v_n \\ f(v_n) &= \lambda v_n. \end{aligned}$$

Hence the matrix of f with respect to $\{v_1, \ldots, v_n\}$ is as claimed. □

Theorem 39. *If the minimal polynomial of f is $(X - \lambda_1)^{m_1} \cdots (X - \lambda_k)^{m_k}$, then an ordered basis of V can be chosen so that the matrix of f is of the form* $\operatorname{diag}(A_1, \ldots, A_k)$ *where each A_i is a square matrix* $\operatorname{diag}(C_{i1} \ldots, C_{ik_i})$ *and where C_{ij} is $J_{n_{ij}}(\lambda_i)$.*

Proof. This follows from Theorem 31 and Lemma 38. □

The matrix of the theorem is called a Jordan canonical form of f. As each square matrix A of size n gives rise to a linear map f of F^n, we define the notions of minimal polynomial, invariant factor and elementary divisors of A as those of f.

It follows that two square matrices over F are similar iff they have the same invariant factors (or they have the same elementary divisors). If $F = \mathbb{C}$ or any algebraically closed field, then every square matrix is similar to a Jordan canonical matrix.

Remark 40. With a little more work, Theorem 25 and Theorem 31 can also be established for finitely generated modules over a principal ideal domain. We refer the reader to the set of notes "A Course in Module Theory" by Amber Habib and Kumaresan as well as any graduate level text on Algebra (such as the ones by Dummit and Foote or by Hungerford) for details. I thank Professors N. Vanaja and M.I. Jinnah for a careful reading of the manuscript and suggestions.