Modules over a PID
Hjalmar Rosengren, 20 October 2015

We will classify all finitely generated modules over a PID R. One very important case is $R = \mathbb{Z}$, when we obtain a classification of finitely generated abelian groups. We will show that any such group is isomorphic to

$$\mathbb{Z}^s \times \mathbb{Z}/p_1^{k_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_m^{k_m}\mathbb{Z}, \tag{1}$$

where p_j are primes. This is a good example to keep in mind throughout the discussion. Another important case gives the "Jordan normal form" of a complex matrix, which generalizes diagonalization to non-diagonalizable matrices. This is useful for many applications of linear algebra.

To formulate the main theorem, consider the equivalence relation on prime elements of R, defined by $p \sim q$ if $p = eq$ with e a unit. Let $\mathfrak{P}$ be a set of representatives for these equivalence classes. (For instance, in $\mathbb{Z}$ the equivalence classes are $\{\pm p\}$ with p a prime number. Choosing always the positive representative, we can take $\mathfrak{P}$ as the set of prime numbers.)

Theorem 1. *Any finitely generated module over a* PID R *is isomorphic to*

$$R^s \times R/p_1^{k_1}R \times \cdots \times R/p_m^{k_m}R \tag{2}$$

with $p_j \in \mathfrak{P}$. *Moreover, this decomposition is unique up to rearranging the factors.*

The decomposition (2) is called the *primary decomposition* of M (an ideal of the form p^kR with p prime is called a primary ideal). Note that some of the primes p_j in Theorem 1 may be equal. As an example, if G is a finite abelian group of order 24, then G is a product as in (1), where the factor $\mathbb{Z}^s$ is absent ($s = 0$) and where $p_1^{k_1} \cdots p_m^{k_m} = 24 = 2^3 \cdot 3$. Up to reordering the factors, the only such groups are

$$\mathbb{Z}_8 \times \mathbb{Z}_3, \qquad \mathbb{Z}_4 \times \mathbb{Z}_2 \times \mathbb{Z}_3, \qquad \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3. \tag{3}$$

The proof of Theorem 1 is rather long. The proof of exixtence of the decomposition (2) will be divided into three steps, which we call A, B and C. To illustrate these steps we consider the example

$$M = \mathbb{Z}^2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3.$$

In step A, we split M as

$$M = \mathbb{Z}^2 \times (\mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3),$$

where the first factor is a free module and the second one a so called torsion module. In step B, the torsion part is split into factors corresponding to distinct primes. In the example,

$$M = \mathbb{Z}^2 \times (\mathbb{Z}_2 \times \mathbb{Z}_4) \times \mathbb{Z}_3.$$

In the final step C, the factor related to each prime p is split into modules of the form R/p^kR; in the example,

$$M = \mathbb{Z}^2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3.$$

Before we start the proof, we note that the individual factors in (1) are cyclic groups, which can either be defined as groups with one generator or as quotients of $\mathbb{Z}$. The corresponding concept for modules is *cyclic modules*.

Lemma 2. *A module over a commutative ring R is generated by one element if and only if it is isomorphic to a quotient R/I for some ideal I. More precisely, a module Rx generated by x is isomorphic to R/I with $I = \mathrm{Ann}(x) = \{r \in R;\, rx = 0\}$ (the annihilator of x).*

This also holds over non-commutative rings, but for left ideals rather than ideals.

Proof. If $M = Rx$, then $\phi(r) = rx$ gives a surjective homomorphism $R \to M$ with kernel $I = \mathrm{Ann}(x)$, so $Rx \simeq R/I$. Conversely, if we let $x = 1 + I \in R/I$, then x generates R/I and $\mathrm{Ann}(x) = I$. □

Step A

We now start with Step A in the proof of Theorem 1. We define a *torsion element* of a module M as an element x such that $ax = 0$ for some non-zero $a \in R$. The set of all torsion elements form a submodule, which we denote M_{tor}. A module with $M_{\mathrm{tor}} = \{0\}$ is called *torsion free* and a module with $M_{\mathrm{tor}} = M$ is called a *torsion module*. It is easy to see that M/M_{tor} is torsion free.

Lemma 3. *A free module M over an integral domain R is torsion-free.*

Proof. Let $(e_j)_{j=\in\Lambda}$ be a basis for M. Suppose that $ax = 0$, where $x = \sum_j x_j e_j$. Then, $\sum_j ax_j e_j = 0$ so by the definition of a basis $ax_j = 0$ for each j. Since R is a domain we have either $a = 0$ or $x_j = 0$ for all j, that is, $x = 0$. □

In the context of (2), Lemma 3 shows that the factor R^s is torsion free. The remaining factors form a torsion module, since they are annihilated by $p_1^{k_1} \cdots p_m^{k_m}$.

We will write $\dim(M)$ for the number of elements in a basis for a module. Recall that this makes sense for finitely generated free modules over commutative rings (but not over rings in general).

Proposition 4. *If F is a finitely generated free module over a* PID *R and M is a submodule, then M is again free and* $\dim(M) \leq \dim(F)$.

The condition of R being a PID is vital. Indeed, consider an integral domain R as a module over itself. Then, R is free (with basis 1). A submodule of R is simply an ideal I. We claim that a non-principal ideal is not free. Indeed, if e_1 and e_2 are two distinct elements from a basis for I, then the equation $e_1 e_2 = e_2 e_1$ expresses the same element of I as a linear combination of basis elements in two distinct ways, which is impossible. Thus, if I has a basis it can consist only of one element, but then I is principal.

Proof of Proposition 4. Let $e_1, \ldots, e_n$ be a basis for F. Let

$$M_k = M \cap (Re_1 + \cdots + Re_k).$$

We will prove by induction on k that M_k is free and $\dim M_k \leq k$. Our starting point is the trivial case $M_0 = \{0\}$ and the endpoint case $k = n$ is the statement of the theorem.

Consider the map $\pi : M_k \to R$ defined by

$$\pi(x_1 e_1 + \cdots + x_k e_k) = x_k.$$

Then π is an R-module homomorphism so $\mathrm{Im}(\pi)$ is an ideal. By the PID property, $\mathrm{Im}(\pi) = aR$ for some $a \in R$. If $a = 0$, clearly $M_k = M_{k-1}$ and we are done.

Assuming from now on that $a \neq 0$, pick $x \in M_k$ with $\pi(x) = a$. We claim that $M_k = M_{k-1} \oplus Rx$ or, equivalently, $M_k = M_{k-1} + Rx$ and $M_{k-1} \cap Rx = \{0\}$. For the first identity, pick any $y \in M_k$. We have $\pi(y) = ab$ for some

$b \in R$. Then, $\pi(y-bx) = 0$, so $y-bx \in M_{k-1}$, which gives $y \in M_{k-1}+Rx$. For the second identity, if $rx \in M_{k-1}$ then $0 = \pi(rx) = r\pi(x) = ra$. Since $a \neq 0$ we get $r = 0$ (R is a domain) and thus $rx = 0$. Finally, we note that Rx is a free module with basis x. Otherwise, we would have $rx = 0$ for some $r \neq 0$. By Lemma 3, that would contradict F being free. Thus, it follows from the induction hypothesis that M_k is free with $\dim(M_k) = \dim(M_{k-1})+1 \leq k$. □

Corollary 5. *Let M be a finitely generated module over a* PID*. Then any submodule of M is finitely generated.*

Proof. Suppose M is generated by $v_1, \ldots, v_n$. Define $f : R^n \to M$ by $f(x_1, \ldots, x_n) = x_1v_1 + \cdots + x_nv_n$. Then, f is a homomorphism. If N is a submodule of M then $f^{-1}(N) = \{x \in R^n;\, f(x) \in N\}$ is a submodule of R^n. By Prop. 4, $f^{-1}(N)$ is finitely generated. Acting by f on a set of generators we obtain a finite set of generators for N. □

By Lemma 3, a free module over an integral domain is torsion free. The converse holds for finitely generated modules over a PID.

Lemma 6. *If M is a finitely generated module over a* PID *and M is torsion free then M is free.*

Proof. Let $v_1, \ldots, v_n$ be generators for M and let $e_1, \ldots, e_k$ be a maximal set of linearly independent elements among these generators. Then, $e_1, \ldots, e_k$ generate a free module $F \subseteq M$. We claim that we can find non-zero elements $a_j \in R$ such that $a_jv_j \in F$. If v_j is one of the generators e_k this is true with $a_j = 1$. Else, $\{v_j, e_1, \ldots, e_k\}$ are linearly dependent, so we can write

$$a_jv_j + x_1e_1 + \cdots + x_ke_k = 0, \qquad a_j, x_1, \ldots, x_k \in R,$$

where not all the coefficients are zero. Since $e_1, \ldots, e_k$ are linearly independent, $a_j \neq 0$. We now let $a = a_1 \cdots a_n$. Since $a_jv_j \in F$ we have that $av_j \in F$ for each j. It follows that $f(x) = ax$ is a homomorphism from M to F. Since M is torsion-free, f is injective. Thus, M is isomorphic to a submodule of a finitely generated free module. The conclusion now follows from Prop. 4. □

We are now ready for the following weak version of Theorem 1. This completes Step A in the proof.

Lemma 7. *If M is a finitely generated module over a* PID*, then $M \simeq R^s \times M_{\text{tor}}$ for some s.*

Proof. As we remarked above, M/M_{tor} is torsion free. Thus, by Lemma 6, it is free. By Lemma 4 from the notes last time, M_{tor} is a direct summand of M, so $M = F \oplus M_{\text{tor}}$ for some submodule F. It follows that $F \simeq M/M_{\text{tor}}$, so F is free. Finally, by Cor. 5, F is finitely generated, so $F \simeq R^s$ for some s. □

Step B

We will now study torsion modules. We first say a few words about greatest common divisors. When a and b are elements of a PID, we write $\gcd(a, b) = c$ for any element c such that $(a, b) = (c)$ (it is determined up to multiplication by a unit). We can obtain c as the product of all prime powers that divide both a and b. Since $c \in (a, b)$ we can write $c = as + bt$ for some $s, t \in R$.

When M is a module over R and $a \in R$, let $M_a = \{x \in M;\, ax = 0\}$. Clearly, M_a is a sub-module.

Lemma 8. *If M is a module over a* PID *R and a, b, $c \in R$ with $a = bc$ and $\gcd(b, c) = 1$, then*

$$M_a = M_b \oplus M_c.$$

Proof. We can write $1 = sc + tb$ for some s, $t \in R$. We need to show that $M_a = M_b + M_c$ and that $M_b \cap M_c = \{0\}$. Both statements follow from writing $x = scx + tbx$. Indeed, if $x \in M_a$, then $scx \in M_b$ and $tbx \in M_c$. Moreover, if $x \in M_b \cap M_c$ then $scx = tbx = 0$. □

We can now obtain a preliminary decomposition of a finitely generated torsion module. This completes Step B in the proof of Theorem 1.

Lemma 9. *If M is a finitely generated torsion module over a* PID*, then*

$$M = M_{p_1^{k_1}} \oplus \cdots \oplus M_{p_m^{k_m}} \tag{4}$$

for certain distinct primes $p_j \in \mathfrak{P}$ and positive integers k_j.

Note that, in contrast to the decomposition (2), the p_j appearing in (4) are distinct. We should think of each summand in (4) as gathering all factors in (2) that involve a fixed prime.

Proof of Lemma 9. The *annihilator* of M is defined as

$$\text{Ann}(M) = \{a \in R;\, ax = 0 \text{ for all } x \in M\}.$$

It is an ideal in R. We prove that $\mathrm{Ann}(M) \neq \{0\}$. Indeed, if $v_1, \ldots, v_n$ are generators for M, then $a_j v_j = 0$ for some non-zero $a_j \in R$ and it follows that $a_1 \cdots a_n \in \mathrm{Ann}(M)$.

By the PID property, we can write $\mathrm{Ann}(M) = aR$ for some non-zero $a \in R$. Since a PID is a UFD, we may factor a as a product of primes. Since multiplying a by a unit does not change the ideal aR, we may assume $a = p_1^{k_1} \cdots p_m^{k_m}$, where $p_j \in \mathfrak{P}$. The desired result now follows by iterating Lemma 8. □

Step C

We will now look at the individual terms in (4). Note that, for any module M, $\mathrm{Ann}(M_{p^k})$ is an ideal in R containing $p^k R$ and thus has the form $p^l R$ for some $l \leq k$. Thus, the following result shows that each summand in (4) can be decomposed as a product of modules of the form $R/p^l R$.

Lemma 10. *Let M be a finitely generated module over a* PID *R, such that $\mathrm{Ann}(M) = p^k R$, where $p \in \mathfrak{P}$ and $k \geq 0$. Then, M is isomorphic to a product*

$$R/(p^{l_1}R) \times \cdots \times R/(p^{l_m}R) \tag{5}$$

for some positive integers l_j.

To prove Lemma 10, we will argue by induction on the "size" of M. It's not immediately obvious how the size should be measured, but it turns out that the following idea works. It's easy to see that, for any module M, M/pM is a module over R/pR (cf. Prop. (3.14) in Brzezinski). (Note that R/pR is a field, so M/pM is in fact a vector space.) Moreover, if M is finitely generated, then the cosets containing the generators generate M/pM. Thus, $|M|_p = \dim_{R/pR}(M/pM)$ is a non-negative integer. We will prove Lemma 10 by induction on $|M|_p$. It is easy to check that if M is given by (5), then $|M|_p = m$ (cf. Lemma 11 below). Thus, we are actually performing induction over the number of factors in (5), but of course that does not make sense until we have proved the lemma.

Proof of Lemma 10. If M is the trivial module, we interpret (5) as an empty product. Otherwise, $k \geq 1$ and there is an element $x \in M$ such that $p^{k-1}x \neq 0$. Note that $x \notin pM$ since $x = py$ would give $p^k y \neq 0$. This means in particular that $pM \subsetneq M$ so that $|M|_p \geq 1$. Thus, $|M|_p = 0$ only for the trivial module, which we can use as the starting case for the induction.

Assume that the statement of the lemma holds for all modules N with $|N|_p < |M|_p$. Choose x as above and let $N = M/Rx$. We claim that $|N|_p < |M|_p$. To see this, note that the natural projection $M \to N/pN$ maps pM to 0, so there is a surjective homomorphism $\phi : M/pM \to N/pN$. As we have observed above, $x \notin pM$, so x is a non-trivial element in $\mathrm{Ker}(\phi)$. By the dimension theorem, it follows that $|M|_p = |N|_p + \dim_{R/pR}(\mathrm{Ker}(\phi)) > |N|_p$.

Having proved that $|N|_p < |M|_p$, it follows from our induction hypothesis that N is isomorphic to a product like (5). If we can show that the sequence

$$0 \to Rx \to M \to N \to 0 \tag{6}$$

splits, then $M \simeq N \times Rx$. Since, by Lemma 2, $Rx \simeq R/p^kR$, this would complete the proof. Using again Lemma 2, if N has the form (5) then we can write

$$N = R\bar{x}_1 \oplus \cdots \oplus R\bar{x}_n, \tag{7}$$

where $\bar{x}_j \in M/Rx$ are such that $\mathrm{Ann}(\bar{x}_j) = p^{l_j}R$. To split the sequence, we need to find representatives y_j for the coset $\bar{x}_j$ so that

$$\psi(a_1\bar{x}_1 + \cdots + a_m\bar{x}_m) = a_1y_1 + \cdots a_my_m \tag{8}$$

is a well-defined homomorphism from N to M. The only potential problem with this definition is that the coefficients $a_j \in R$ are not uniquely determined by $\sum a_j\bar{x}_j$. If $\sum a_j\bar{x}_j = \sum_j b_j\bar{x}_j$ then, since the sum (7) is direct, $b_j - a_j \in \mathrm{Ann}(\bar{x}_j) = p^{l_j}R$ for each j and thus $b_j = a_j + p^{l_j}r_j$ for some $r_j \in R$. The right-hand side of (8) is invariant under $a_j \mapsto b_j$ provided that $p^{l_j}y_j = 0$. If we can find such representatives y_j for $\bar{x}_j$, then ψ splits the sequence and the proof is complete.

To find appropriate representatives y_j, let x_j be an arbitrary representative of $\bar{x}_j$. Then $p^{l_j}x_j \in Rx$, so we can write $p^{l_j}x_j = p^scx$, where $(c,p) = 1$. Since $p^kx = 0$, we may assume $s \le k$. If $s = k$ we can choose $y_j = x_j$. If $s \le k-1$, then $p^{k-1-s+l_j}x_j = p^{k-1}cx$. We want to prove that $p^{k-1}cx \neq 0$. To this end, we write $1 = tp+uc$, which gives $0 \neq p^{k-1}x = tp^kx+up^{k-1}cx = up^{k-1}cx$. It follows that $k-1-s+l_j \le k-1$, that is, $s \ge l_j$. With $y_j = x_j - p^{s-l_j}cx$, we then have $y_j \in x_j + Rx$ and $p^{l_j}y_j = 0$. As we have seen, with this choice of y_j the map (8) splits the sequence (6) and the proof is complete. □

Together, Lemma 7, Lemma 9 and Lemma 10 prove the existence part of Theorem 1.

For the uniqueness, the following simple result will be useful.

Lemma 11. *If R is a* PID*, p and q are prime elements and $M = R/q^kR$, then*

$$p^lM/p^{l+1}M \simeq \begin{cases} R/pR, & p \sim q \text{ and } l \leq k-1, \\ 0, & \text{else} \end{cases}$$

as R-modules (and hence also as R/pR-modules).

Proof. By Lemma 3, we can write $M = Rx$ with $\mathrm{Ann}(x) = q^kR$. Then, p^lM is generated by p^lx and $N = p^lM/p^{l+1}M$ is generated by the coset $y = p^lx + p^{l+1}M$. Clearly $py = 0$ so, again by Lemma 3, we have either $N \simeq R/pR$ (when $p^lx \notin p^{l+1}M$ so that $y \neq 0$) or $N = \{0\}$ (when $p^lx \in p^{l+1}M$). If $p = q$ and $l \geq k$, then $p^lx = 0$ so we are in the second case. If $p = q$ and $l \leq k-1$, then $p^lx = p^{l+1}z$ would give $p^{k-1}x = p^{k-1-l}p^lx = p^kz = 0$, which contradicts $\mathrm{Ann}(x) = p^kR$. Then we are in the first case. Finally, if $p \not\sim q$ then $1 = tq^k + up^{l+1}$ for some t and u. Then, $z = up^{l+1}z \in p^{l+1}M$ for any $z \in M$, so we are in the second case. □

We can now prove the uniqueness part of Theorem 1. We need to show that if M denotes the module (2), then s, p_j and k_j can be constructed uniquely from M (up to reordering). It is clear that $s = \dim_R(M/M^{\mathrm{tor}})$ so it's enough to consider the torsion part. If M is a torsion module (that is, there is no factor R^s in (2)) let us compute the numbers

$$d_l(p) = \dim_{R/pR}(p^lM/p^{l+1}M) \tag{9}$$

for $p \in \mathfrak{P}$ and $l \geq 0$. By generalities on direct products, we can compute $d_l(p)$ by adding up the contribution from each factor. Thus, by Lemma 11, $d_l(p)$ is the number of indices j such that $p_j = p$ and $l \leq k_j - 1$. Consequently,

$$d_{l-1}(p) - d_l(p) \tag{10}$$

is the number of indices j such that $p_j = p$ and $k_j = l$. This shows that p_j and k_j can be reconstructed from the module M. The proof of Theorem 1 is now complete.

It is sometimes useful to rewrite the decomposition (2) in the following alternative form.

Theorem 12. *Any finitely generated module M over a* PID *R is isomorphic to*

$$R^s \times R/q_1R \times R/q_2R \times \cdots \times R/q_mR, \tag{11}$$

where q_j are non-zero non-units of R and $q_1|q_2|\ldots|q_m$. Moreover, this decomposition is unique (up to multiplying q_j by units).

The elements q_j are called *invariant factors* and (11) the *invariant factor decomposition*. In particular, a finitely generated torsion module is isomorphic to (11), where the factor R^s is absent ($s = 0$). In this case, $\mathrm{Ann}(M) = q_m R$.

To prove the existence part of Theorem 12, we write M as in (2). We may assume that $s = 0$. We then rearrange the product in a rectangular array so that all factors involving a fixed prime are placed in the same row, with the entries in each row ordered by increasing exponent of p. Moreover, the rows are aligned to the right so that the right-most entry of each row ends up in the same column. We then define q_k as the product of all prime powers in the k-th column.

An example should clarify the definition of q_k. If

$$R = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/16\mathbb{Z} \times \mathbb{Z}/25\mathbb{Z},$$

we write

$$\begin{array}{lll} R = \mathbb{Z}/2\mathbb{Z} & \times \mathbb{Z}/2\mathbb{Z} & \times\mathbb{Z}/16\mathbb{Z} \\ & & \times\mathbb{Z}/3\mathbb{Z} \\ & \times \mathbb{Z}/5\mathbb{Z} & \times\mathbb{Z}/25\mathbb{Z}. \end{array}$$

Then, $q_1 = 2$, $q_2 = 2 \cdot 5$ and $q_3 = 16 \cdot 3 \cdot 25$.

Having defined q_k in this way, it is obvious that $q_k \mid q_{k+1}$. Moreover, by (12) below, each $R/q_k R$ is isomorphic to the direct product of those factors in (2) that were put in the k-th column (in the example above, we have for instance that $\mathbb{Z}/10\mathbb{Z} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}$). This shows the existence part.

To prove uniqueness, we reverse the argument. Given a factorization (11) with $q_j \mid q_{j+1}$ we can factorize q_j into primes and construct a rectangular array of factors $R/p^k R$ as above. This leads to a decomposition of M as in (2). Since that factorization is unique, so is the factorization (11).

To prove (12), we use the module case of the following fact.

Lemma 13. *If R is a* PID *and $a, b \in R$ with $\gcd(a, b) = 1$, then*

$$R/abR \simeq R/aR \times R/bR$$

as rings, and also as R-modules.

Proof. Consider the map $f : R \to R/aR \times R/bR$ given by $f(x) = (\bar{x}, \bar{x})$. Clearly, $\mathrm{Ker}(f) = aR \cap bR = abR$. To show that f is surjective, take $(\bar{x}_1, \bar{x}_2)$ arbitrary and write $1 = sa + tb$. Then, if we let

$$x = x_1 + sa(x_2 - x_1) = x_2 + tb(x_1 - x_2)$$

we have $f(x) = (\bar{x}_1, \bar{x}_2)$. Note that f is both a ring homomorphism and an R-module homomorphism, since

$$f(xy) = (\overline{xy}, \overline{xy}) = (\bar{x}, \bar{x})(\bar{y}, \bar{y}) = f(x)f(y)$$

in the ring and

$$f(x, y) = (\overline{xy}, \overline{xy}) = x(\bar{y}, \bar{y}) = xf(y)$$

in the module. □

Iterating this lemma, we have that if $a = p_1^{k_1} \cdots p_m^{k_m}$, with p_j distinct prime elements, then

$$R/aR \simeq R/p_1^{k_1}R \times \cdots \times R/p_m^{k_m}R, \tag{12}$$

both as rings and as R-modules.

Canonical forms

We now come to an important application of the theory described above, canonical forms for matrices. Any linear operator on a finite-dimensional vector space can be represented by a matrix, which expresses how it acts on a basis. However, changing the basis leads to a different matrix. We would like to find a basis such that the matrix has an especially simple form. Moreover, the form should be canonical in the sense that this basis is unique (possibly up to reordering the basis vectors). As a conseqence, if we want to know whether two matrices are *similar* in the sense that they express the same linear map in different bases, we can check whether their canonical forms agree or not. One familiar example is diagonalization, but not any matrix can be diagonalized. The Jordan canonical form explained below can be viewed as an analogue of diagonalization for arbitrary complex matrices.

Let K be a field, V an n-dimensional vector space over K and $A \in \mathrm{End}_K(V)$, that is, A is a linear map from V to itself. The main idea is to study A by viewing V as a module over $R = K[x]$. The module structure is obtained from the ring homomorphism $\Phi : K[x] \to \mathrm{End}_K(V)$ given by $\Phi(p) = p(A)$. More explicitly,

$$(k_0+k_1x+\cdots+k_mx^m)v = k_0v+k_1Av+k_2A^2v+\cdots+k_mA^mv, \quad k_i \in K, \quad v \in V.$$

It is natural to ask what propertied of A are reflected by the module structure of V. The following simple lemma gives a very clear answer to that question.

Lemma 14. *If V_A denotes the $K[x]$-module associated to $A \in \mathrm{End}_K(V)$, then $V_A \simeq V_B$ if and only if A and B are similar over K, that is, $A = T^{-1}BT$ for some invertible element $T \in \mathrm{End}_K(V)$.*

The proof is trivial: a $K[x]$-module isomorphism $T : V_A \to V_B$ is an invertible element of $\mathrm{End}_K(V)$ such that $p(A) = T^{-1}p(B)T$ for all polynomials p. With $p(x) = x$, this gives $A = T^{-1}BT$. Conversely, if $A = T^{-1}BT$, then $A^k = T^{-1}B^kT$ and consequently $p(A) = T^{-1}p(B)T$ for all polynomials p.

Let us now start investigating V as a $K[x]$-module. We know that R is a PID and V is finitely generated (any basis for V generates V). Note also that, as vector spaces over K, $K[x]$ is infinite-dimensional and $\mathrm{End}_K(V)$ is n^2-dimensional. Since Φ is K-linear, it follows that $\mathrm{Ker}(\Phi) = \mathrm{Ann}(M)$ is non-zero. In particular, V is a torsion module.

We can now apply Thm. 1 and Thm. 12. We have

$$V \simeq R/q_1R \times \cdots \times R/q_mR \tag{13}$$

for some polynomials q_j. We may either choose $q_j = p_j^{k_j}$ with p_j irreducible polynomials or q_j as non-constant polynomials with $q_1 \mid q_2 \cdots \mid q_m$. Note that the units in R are the non-zero constant polynomials, so if we normalize q_j to be monic (leading coefficient 1) then they are unique (in the first case up to reordering).

If ϕ is an R-module isomorphism from V to the right-hand side of (13), then $\phi(Av) = x\phi(v)$. So to understand how A acts on the left we must understand how multiplication by x acts on the right. Consider first the action on R/qR, where $q(x) = a_0 + a_1x + \cdots + a_{k-1}x^{k-1} + x^k$ is a monic polynomial of degree k. We choose 1, x, $x^2, \ldots, x^{k-1}$ as a basis for R/qR. Then, multiplication by x maps each basis element to the next, except that x^{k-1} is mapped to $x^k = -a_0 - a_1x - \cdots - a_{k-1}x^{k-1}$. Thus, the matrix for multiplication by x is given by

$$M_q = \begin{bmatrix} 0 & 0 & \cdots & -a_0 \\ 1 & 0 & & -a_1 \\ 0 & 1 & & -a_2 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & -a_{k-1} \end{bmatrix}, \tag{14}$$

the so called *companion matrix* to q. We can then interpret (13) as saying that there is a basis for V where A is expressed by the block matrix

$$\begin{bmatrix} M_{q_1} & 0 & \cdots & 0 \\ 0 & M_{q_2} & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & M_{q_m} \end{bmatrix}. \tag{15}$$

In the first case ($q_j = p_j^{k_j}$), we call (15) the *primary canonical form* and in the second case the *rational canonical form*. (Brzezinski calls the first of these "rational canonical form", which seems quite unorthodox.) The primary canonical form is unique up to reordering the blocks whereas the rational canonical form is unique. We remark that, by expanding (14) along the right column, it is easy to see that $\det(xI - M_q) = q(x)$. Thus, the characteristic polynomial for A is in both cases given by $\det(xI - A) = q_1 \cdots q_m$.

One important difference between these two forms is that the primary canonical form depends heavily on the base field. Given, say, a matrix with integer entries, the primary canonical form will typically look different if we work over $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. This is because these fields have different irreducible polyomials: x^2+1 is irreducible over $\mathbb{R}$ but not over $\mathbb{C}$ and x^2-2 is irreducible over $\mathbb{Q}$ but not over $\mathbb{R}$. By contrast, the rational canonical form does not change if we extend the field. This follows from uniqueness; if $K \subseteq K'$ the canonical form over K is also a canonical form over K'. One non-trivial consequence is that the notion of similarity does not depend on the field. For instance, if A and B are integer matrices such that $A = TBT^{-1}$ for a complex matrix T, then there is such a matrix T with rational entries.

Let us look more closely at the primary decomposition in the case when K is algebraically closed. (In fact, it is enough to assume that K contains all eigenvalues of A.) For instance, we can take $K = \mathbb{C}$. Then, any monic irreducible polynomial has the form $x - \lambda$ for some $\lambda \in K$. Consider the matrix for multiplication by x in the module R/qR, with $q = (x - \lambda)^k$. Instead of choosing the basis vectors x^j as above, it's nicer to work with $(x - \lambda)^j$. Then, $x(x - \lambda)^j = \lambda(x - \lambda)^j + (x - \lambda)^{j+1}$, so we obtain instead of M_q the matrix

$$J_q = \begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 1 & \lambda & 0 & & 0 \\ 0 & 1 & \lambda & & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}. \tag{16}$$

In the corresponding basis, A takes the form

$$\begin{bmatrix} J_{q_1} & 0 & \cdots & 0 \\ 0 & J_{q_2} & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & J_{q_m} \end{bmatrix}. \tag{17}$$

This is called the *Jordan canonical form*. It is unique up to reordering the blocks.

In all but the very simplest cases, it makes little sense to compute canonical forms by hand. Therefore, we will not focus on how that can be done, except for a brief discussion on the case $K = \mathbb{C}$. Note that $\mathbb{C}[x]/(x-\lambda) \simeq \mathbb{C}$,

so with $p = x - \lambda$ the numbers (9) are given by

$$\begin{aligned} d_l &= \operatorname{rank}((A-\lambda I)^l) - \operatorname{rank}((A-\lambda I)^{l+1}) \\ &= \dim \operatorname{Ker}((A-\lambda I)^{l+1})) - \dim \operatorname{Ker}((A-\lambda I)^l)). \end{aligned}$$

It is easy to understand directly why the primary canonical form is determined by the solutions to $(A-\lambda I)^l v = 0$. If J is the block (16), then we have e.g.

$$(J-\lambda)^2 = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & \\ 0 & 0 & 0 & & 0 & \\ 1 & 0 & 0 & & 0 & \\ 0 & 1 & 0 & & 0 & \\ 0 & 0 & 1 & & 0 & \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \end{bmatrix}.$$

In general, $(J-\lambda)^k$ has only zeroes except for a diagonal of 1:s, which is pushed to the southwest as k increases. Thus, the solutions of the equation $(J-\lambda)^k v = 0$ is the linear span of the k right-most columns. For the block matrix (17), the solution space is spanned by the k right-most columns in all blocks with λ on the diagonal.

To give a simple example, suppose we are given a complex 6×6-matrix A. We compute the characteristic polynomial $\det(xI - A) = (x-1)^4(x-2)^2$. We start with the eigenvalue $x = 1$ and look for eigenvectors. It turns out that the equation $(A - I)v = 0$ has a two-dimensional space of solutions. This means that there are exactly two Jordan blocks with diagonal entries 1. The sum of their dimension must be equal to 4. So the blocks have either size $(3, 1)$ or $(2, 2)$. Next, we look at the equation $(A-I)^2 v = 0$. In the first case, the solution space would be 3-dimensional and in the second case 4-dimensional. Let's say that we are in the first case. We then turn to the eigenvalue 2 and find that the space of eigenvectors is two-dimensional. There are then two blocks with diagonal entries 2, which necessarily have size 1. We conclude that the Jordan canonical form for A is

$$\begin{bmatrix} 1 & & & & & \\ 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & & 1 & & \\ & & & & 2 & \\ & & & & & 2 \end{bmatrix},$$

where we didn't write the zero entries. The prime powers corresponding to the Jordan blocks are $(x-1)^3 = x^3 - 3x^2 + 3x - 1$, $(x-1)$, $(x-2)$ and $(x-2)$, so the primary canonical form is

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

To get the rational canonical form we arrange the prime powers in an array as explained in the proof of Theorem 2. We get

$$\begin{array}{cc} (x-1) & (x-1)^3 \\ (x-2) & (x-2) \end{array}$$

and can read of the invariant factors as the columns: $c_1 = (x-1)(x-2)$ and $c_2 = (x-1)^3(x-2)$. Note that $c_1 c_2$ is indeed the characteristic polynomial. Since

$$(x-1)^3(x-2) = x^4 - 5x^3 + 9x^2 - 7x + 2,$$

the rational canonical form is

$$\begin{bmatrix} & -2 & & & & \\ 1 & 3 & & & & \\ & & & & & -2 \\ & & & 1 & & 7 \\ & & & & 1 & -9 \\ & & & & & 5 \end{bmatrix}.$$

As a final remark, note that since $\mathrm{Ann}(M)$ is a non-zero ideal in R, it is equal to qR for a unique monic polynomial q, called the *minimal polynomial* of A. Equivalently, q is the smallest degree monic polynomial such that $q(A) = 0$. In terms of the rational canonical form, it is clear that $q = q_m$, the largest invariant factor of M. As was remarked above, the characteristic polynomial for A is $p = q_1 \cdots q_m$. This implies the following useful result.

Theorem 15 (Cayley–Hamilton)**.** *If p is the characteristic polynomial of a square matrix A, then $p(A) = 0$.*