M.6. Rational canonical form

In this section we apply the theory of finitely generated modules of a principal ideal domain to study the structure of a linear transformation of a finite dimensional vector space.

If T is a linear transformation of a finite dimensional vector space V over a field K, then V has a $K[x]$–module structure determined by $f(x)v = f(T)v$ for $f(x) \in K[x]$ and $v \in V$. Since V is finitely generated as a K–module, it is finitely generated as a $K[x]$–module. Moreover, V is a torsion module over $K[x]$. In fact, if V is n–dimensional, then $\mathrm{End}_K(V)$ is an n^2–dimensional vector space over K, so the n^2+1 linear transformations $\mathrm{id}, T, T^2, \dots, T^{n^2}$ are not linearly independent. Therefore, there exist $\alpha_0, \dots, \alpha_{n^2}$ such that $\sum_{j=0}^{n^2} \alpha_j T^j = 0$ in $\mathrm{End}_K(V)$. But this means that the polynomial $\sum_{j=0}^{n^2} \alpha_j x^j$ is in the annihilator of V in $K[x]$.

A $K[x]$–submodule of V is a vector subspace V_1 that is invariant under T, $Tv \in V_1$ for all $v \in V_1$. If $(x_1, \dots, x_n)$ is an ordered basis of V such that the first k basis elements form a basis of V_1, then the matrix of T with respect to this basis has the block triangular form:

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

If $V = V_1 \oplus V_2$ where both V_1 and V_2 are invariant under T, and $(x_1, \dots, x_n)$ is an ordered basis of V such that the first k elements constitute a basis of V_1 and the remaining elements constitute a basis of V_2, then the matrix of T with respect to this basis has the block diagonal form:

$$\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}.$$

If V is the direct sum of several T–invariant subspaces,

$$V = V_1 \oplus \cdots \oplus V_s,$$

then with respect to an ordered basis that is the union of bases of the subspaces V_i, the matrix of T has the block diagonal form:

$$\begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & A_s \end{bmatrix}.$$

In this situation, let T_i denote the restriction of T to the invariant subspace subspace V_i. In the block diagonal matrix above, A_i is the matrix of T_i with respect to some basis of V_i. We write

$$(T, V) = (T_1, V_1) \oplus \cdots \oplus (T_s, V_s),$$

or just

$$T = T_1 \oplus \cdots \oplus T_s$$

to indicate that V is the direct sum of T–invariant subspaces and that T_i is the restriction of T to the invariant subspace V_i.

A strategy for understanding the structure of a linear transformation T is to find such a direct sum decomposition so that the component transformations T_i have a simple form.

Because V is a finitely generated torsion module over the Euclidean domain $K[x]$, according to Theorem M.5.2, (T, V) has a direct sum decomposition

$$(T, V) = (T_1, V_1) \oplus \cdots \oplus (T_s, V_s),$$

where V_i is a cyclic $K[x]$–module

$$V_i \cong K[x]/(a_i(x)),$$

$\deg(a_i(x)) \geq 1$ (that is, $a_i(x)$ is not zero and not a unit) and $a_i(x)$ divides $a_j(x)$ if $i \leq j$. Moreover, if we insist that the $a_i(x)$ are monic, then they are unique. We call the polynomials $a_i(x)$ the *invariant factors* of T.

To understand the structure of T, it suffices to understand how T_i acts on the cyclic $K[x]$–module V_i.

Definition M.6.1. The *companion matrix* of a monic polynomial $a(x) = x^d + \alpha_{d-1}x^{d-1} + \cdots + \alpha_1 x + \alpha_0$ is the matrix

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & 0 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & -\alpha_{d-2} \\ 0 & 0 & 0 & \cdots & 1 & -\alpha_{d-1} \end{bmatrix}$$

We denote the companion matrix of $a(x)$ by C_a.

Lemma M.6.2. *Let T be a linear transformation on a finite dimensional vector space V over K and let*

$$a(x) = x^d + \alpha_{d-1}x^{d-1} + \cdots + \alpha_1 x + \alpha_0 \in K[x].$$

The following conditions are equivalent:

(a) *V is a cyclic $K[x]$–module with annihilator ideal generated by $a(x)$.*

(b) *V has a vector v_0 such that $V = \mathrm{span}(\{T^j v_0 : j \geq 0\})$ and $a(x)$ is the monic polynomial of least degree such that $a(T)v_0 = 0$.*

(c) *$V \cong K[x]/(a(x))$ as $K[x]$ modules.*

(d) *V has a basis with respect to which the matrix of T is the companion matrix of $a(x)$.*

Proof. We already know the equivalence of (a)-(c), at least implicitly, but let us nevertheless prove the equivalence of all four conditions. V is a cyclic module with generator v_0, if, and only if, $V = K[x]v_0 = \{f(T)v_0 : f(x) \in K[x]\} = \mathrm{span}\{T^j v_0 : j \geq 0\}$. Moreover, $\mathrm{ann}(V) = \mathrm{ann}(v_0)$ is the principal ideal generated by its monic element of least degree, so $\mathrm{ann}(V) = (a(x))$ if, and only if, $a(x)$ is the polynomial of least degree such that $a(T)v_0 = 0$. Thus conditions (a) and (b) are equivalent.

If (b) holds, then $f(x) \mapsto f(x)v_0$ is a surjective module homomorphism from $K[x]$ to V, and $a(x)$ is an element of least degree in the kernel of this map, so generates the kernel. Hence $V \cong K[x]/(a(x))$ by the homomorphism theorem for modules.

In proving that (c) implies (d), we may assume that V *is* the $K[x]$–module $K[x]/(a(x))$, and that T is the linear transformation

$$f(x) + (a(x)) \mapsto xf(x) + (a(x)).$$

Write $J = (a(x))$ for convenience. I claim that

$$B = (1 + J, x + J, \dots, x^{d-1} + J)$$

is a basis of $K[x]/(a(x))$ over K. In fact, for any $f(x) \in K[x]$, we can write $f(x) = q(x)a(x) + r(x)$ where $r(x) = 0$ or $\deg(r(x)) < d$. Then $f(x) + J = r(x) + J$, which means that B spans $K[x]/(a(x))$ over K. If B is not linearly independent, then there exists a nonzero polynomial $r(x)$ of degree less than d such that $r(x) \in J$; but this is impossible since $J = (a(x))$. The matrix of T with respect to B is clearly the companion matrix of $a(x)$, as $T(x^j + J) = x^{j+1} + J$ for $j \leq d-2$ and $T(x^{d-1}+J) = x^d + J = -(a_0 + a_1x + \cdots + a_{d-1}x^{d-1}) + J$.

Finally, if V has a basis $B = (v_0, \dots, v_{d-1})$ with respect to which the matrix of T is the companion matrix of $a(x)$, then $v_j = T^j v_0$ for $j \leq d-1$ and $T^d v_0 = Tv_{d-1} = -(\sum_{i=0}^{d-1} \alpha_i v_i) = -(\sum_{i=0}^{d-1} \alpha_i T^i)v_0$. Therefore, V is cyclic with generator v_0 and $a(x) \in \mathrm{ann}(v_0)$. No polynomial of degree less than d annihilates v_0, since $\{T^j v_0 : j \leq d-1\} = B$ is linearly independent. This shows that condition (d) implies (b). ■

Definition M.6.3. Say that a matrix is in *rational canonical form* if it is block diagonal

$$\begin{bmatrix} C_{a_1} & 0 & \cdots & 0 \\ 0 & C_{a_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & C_{a_s} \end{bmatrix},$$

where C_{a_i} is the companion matrix of a monic polynomial $a_i(x)$ of degree ≥ 1, and $a_i(x)$ divides $a_j(x)$ for $i \leq j$

Theorem M.6.4. *(Rational canonical form) Let T be a linear transformation of a finite dimensional vector space V over a field K.*

(a) *There is an ordered basis of V with respect to which the matrix of T is in rational canonical form.*

(b) *Only one matrix in rational canonical form appears as the matrix of T with respect to some ordered basis of V.*

Proof. According to Theorem M.5.2, (T, V) has a direct sum decomposition

$$(T, V) = (T_1, V_1) \oplus \cdots \oplus (T_s, V_s),$$

where V_i is a cyclic $K[x]$–module

$$V_i \cong K[x]/(a_i(x)),$$

and the polynomials $a_i(x)$ are the invariant factors of T. By Lemma M.6.2, there is a basis of V_i such that the matrix of T_i with respect to this basis is the companion matrix of $a_i(x)$. Therefore, there is a basis of V with respect to which the matrix of T is in rational canonical form.

Now suppose that the matrix A of T with respect to some basis is in rational canonical form, with blocks C_{a_i} for $1 \leq i \leq s$. It follows that (T, V) has a direct sum decomposition

$$(T, V) = (T_1, V_1) \oplus \cdots \oplus (T_s, V_s),$$

where the matrix of T_i with respect to some basis of V_i is C_{a_i}. By Lemma M.6.2, $V_i \cong K[x]/(a_i(x))$ as $K[x]$–modules. Thus

$$V \cong K[x]/(a_1(x)) \oplus \cdots \oplus K[x]/(a_s(x)).$$

By the uniqueness of the invariant factor decomposition of V (Theorem xxx), the polynomials $a_i(x)$ are the invariant factors of the

$K[x]$–module V, that is, the invariant factors of T. Thus the polynomials $a_i(x)$, and therefore the matrix A is uniquely determined by T. ■

The matrix in rational canonical form whose blocks are the companion matrices of the invariant factors of T is called *the rational canonical form of* T.

Recall that two linear transformations T_1 and T_2 in $\mathrm{End}_K(V)$ are said to be *similar* if there is an invertible $U \in \mathrm{End}_K(V)$ such that $T_2 = UT_1U^{-1}$. Likewise two matrices A_1 and A_2 are *similar* if there is an invertible matrix S such that $A_2 = SA_1S^{-1}$.

According to the following result, the rational canonical form is a complete invariant for similarity of linear transformations. We will see later that the rational canonical form is computable, so we can actually check whether two transformations are similar by computations.

Proposition M.6.5. *Two linear transformations T_1 and T_2 of a finite dimensional vector space V are similar if, and only if, they have the same rational canonical form.*

Proof. The rational canonical form of a linear transformation T determines, and is determined by, the invariant factor decomposition of the $K[x]$–module corresponding to T, as is clear from the proof of Theorem M.6.4. Moreover, two finitely generated torsion $K[x]$–modules have the same invariant factor decomposition if, and only if, they are isomorphic. So are assertion is equivalent to the statement that T_1 and T_2 are similar if, and only if, the $K[x]$–modules determined by these linear transformations are isomorphic as $K[x]$–modules.

Let V_1 denote V endowed with the $K[x]$–module structure derived from T_1 and let V_2 denote V endowed with the $K[x]$–module structure derived from T_2. Suppose $U : V_1 \longrightarrow V_2$ is a $K[x]$–module isomorphism; then U is a vector space isomorphism satisfying $T_2(Uv) = x(Uv) = U(xv) = U(T_1v)$. It follows that $T_2 = UT_1U^{-1}$.

Conversely, suppose that U is an invertible linear transformation such that $T_2 = UT_1U^{-1}$. It follows that for all $f(x) \in K[x]$, $f(T_2) = Uf(T_1)U^{-1}$; equivalently, $f(T_2)Uv = Uf(T_1)v$ for all $v \in V$ But this means that U is a $K[x]$–module isomorphism from V_1 to V_2. ■

Rational canonical form for matrices

Let A be an n–by–n matrix over a field K. Let T be the linear transformation of K^n determined by left multiplication by A, $T(v) = Av$ for $v \in K^n$. A is the matrix of T with respect to the standard basis of K^n. A second matrix A' is similar to A if, and only if, A' is the matrix of T with respect to some other ordered basis. Exactly one such matrix is in rational canonical form, according to Theorem M.6.4. So we have the following result:

Proposition M.6.6. *Any n–by–n matrix is similar to a unique matrix in rational canonical form.*

Definition M.6.7. The unique matrix in rational canonical form that is similar to a given matrix A is called the *rational canonical form of A*.

The blocks of the rational canonical form of A are companion matrices of monic polynomials $a_1(x), \dots, a_s(x)$ such that $a_i(x)$ divides $a_j(x)$ if $i \le j$. These are called the *invariant factors of* A.

The rational canonical form is a complete invariant for similarity of matrices.

Proposition M.6.8. *Two n–by–n matrices are similar in $\mathrm{Mat}_n(K)$ if, and only if, they have the same rational canonical form.*

Proof. There is exactly one matrix in rational canonical form in each similarity equivalence class, and that matrix is the rational canonical form of every matrix in the similarity class. If two matrices have the same rational canonical form A, then they are both similar to A and therefore similar to each other. ■

Corollary M.6.9. *Suppose $K \subseteq F$ are two fields and A, B are two matrices in $\mathrm{Mat}_n(K)$.*

(a) *The rational canonical form of A in $\mathrm{Mat}_n(F)$ is the same as the rational canonical form of A in $\mathrm{Mat}_n(K)$.*

(b) *A and B are similar in $\mathrm{Mat}_n(F)$ if, and only if, they are similar in $\mathrm{Mat}_n(K)$.*

Proof. The similarity class (or orbit) of A in $\mathrm{Mat}_n(K)$ is contained in the similarity orbit of A in $\mathrm{Mat}_n(F)$, and each orbit contains exactly one matrix in rational canonical form. Therefore, the rational canonical form of A in $\mathrm{Mat}_n(F)$ must coincide with the rational canonical form in $\mathrm{Mat}_n(K)$.

If A and B are similar in $\mathrm{Mat}_n(K)$, they are clearly similar in $\mathrm{Mat}_n(F)$. Conversely, if they are similar in $\mathrm{Mat}_n(F)$, then they have the same rational canonical form in $\mathrm{Mat}_n(F)$. By part (a), they have the same rational canonical form in $\mathrm{Mat}_n(K)$, and therefore they are similar in $\mathrm{Mat}_n(K)$. ■

Computing the rational canonical form

We will now investigate how to actually compute the rational canonical form. Let T be a linear transformation of an n–dimensional vector space with basis $\{e_1, \dots, e_n\}$. Let $A = (a_{i,j})$ be the matrix of T with respect to this basis, so $Te_j = \sum_i a_{i,j} e_i$.

Let F be the free $K[x]$–module with basis $\{f_1, \dots, f_n\}$ and define $\Phi : F \longrightarrow V$ by $\sum_j h_i(x) f_i \mapsto \sum_j h_i(T) e_i$. Then Φ is a surjective $K[x]$–module homomorphism. We need to find the kernel of Φ.

The transformation T can be "lifted" to a $K[x]$–module homomorphism of F by using the matrix A. Define $T : F \longrightarrow F$ by requiring that $Tf_j = \sum_i a_{i,j} f_i$. Then we have $\Phi(Tf) = T\Phi(f)$ for all $f \in F$.

I claim that the kernel of Φ is the range of $x - T$. This follows from three observations:

1. $\mathrm{range}(x - T) \subseteq \ker(\Phi)$.
2. $\mathrm{range}(x - T) + F_0 = F$, where F_0 denotes the set of K–linear combinations of $\{f_1, \dots, f_n\}$.
3. $\ker(\Phi) \cap F_0 = \{0\}$.

The first of these statements is clear since $\Phi(xf) = \Phi(Tf) = T\Phi(f)$ for all $f \in F$. For the second statement, note that for any $h(x) \in K[x]$,

$$h(x)f_j = (h(x) - h(T))f_j + h(T)f_j.$$

Since multiplication by x and application of T commute, there is a polynomial g of two variables such that $h(x) - h(T) = (x - T)g(x, T)$. See Exercise M.6.1 Therefore,

$$(h(x) - h(T))f_j \in \mathrm{range}(x - T),$$

while $h(T)f_j \in F_0$. Finally, if $\sum_i \alpha_i f_i \in \ker(\Phi) \cap F_0$, then $0 = \Phi(\sum_i \alpha_i f_i) = \sum_i \alpha_i e_i$. Hence $\alpha_i = 0$ for all i.

Set $w_j = (x - T)f_j = xf_j - \sum_i a_{i,j} f_i$. I claim that $\{w_1, \dots, w_n\}$ is a basis over $K[x]$ of $\mathrm{range}(x - T) = \ker(\Phi)$. In fact, this set clearly spans range$(x-T)$ over $K[x]$ because $x-T$ is a $K[x]$–module homomorphism. We have

$$[w_1, \dots, w_n] = [f_1, \dots, f_n](xE_n - A), \tag{M.6.1}$$

and the determinant of the matrix $xE_n - A$ is a monic polynomial of degree n in $K[x]$, so in particular nonzero. The matrix $xE_n - A$ is not invertible in $\mathrm{Mat}_n(K[x])$, but it is invertible in $\mathrm{Mat}_n(K(x))$, matrices over the field of rational functions, and this suffices to imply that $\{w_1, \dots, w_n\}$ is linearly independent over $K[x]$. See Exercise M.6.2.

Computing the rational canonical form of T is virtually the same thing as computing the invariant factor decomposition of the $K[x]$–module V derived from T. We now have the ingredients to do this: we have a free module F and a $K[x]$–module homomorphism of F onto V. We have a basis of $\ker(\Phi)$ and the "transition matrix" from a basis of F to the basis of $\ker(\Phi)$, as displayed in Equation (M.6.1). So to compute the invariant factor decomposition, we have to diagonalize the matrix $xE_n - A \in \mathrm{Mat}_n(K[x])$ by row and column operations. We want the diagonal entries of the resulting matrix to be monic polynomials, but this only requires some additional row operations of type two (multiplying a row by unit in $K[x]$.) We can compute invertible matrices P and Q such that

$$P(xE_n - A)Q = D(x) = \mathrm{diag}(1, 1, \dots, 1, a_1(x), a_2(x), \dots, a_s(x)),$$

where the $a_i(x)$ are monic and $a_i(x)$ divides $a_j(x)$ for $i \le j$. The polynomials $a_i(x)$ are the invariant factors of T, so they are all we need in order to write down the rational canonical form of T. But we can actually compute a basis of V with respect to which the matrix of T is in rational canonical form.

We have $xE_n - A = P^{-1}D(x)Q^{-1}$, so

$$[w_1, \dots, w_n]Q^{-1} = [f_1, \dots, f_n]P^{-1}D(x).$$

(Let us mention here that we compute the matrix P as a product of elementary matrices implementing the row operations; we can compute the inverse of each of these matrices without additional effort, and thus we can compute P^{-1} without additional effort.) Set

$$[f_1, \dots, f_n]P^{-1} = [y_1, \dots, y_{n-s}, z_1, \dots, z_s].$$

This is a basis of F over $K[x]$, and

$$[y_1, \dots, y_{n-s}, z_1, \dots, z_s]D(x) = [y_1, \dots, y_{n-s}, a_1(x)z_1, \dots, a_s(x)z_s]$$

is a basis of $\ker(\Phi)$. It follows that

$$\{v_1, \dots, v_s\} := \{\Phi(z_1), \dots, \Phi(z_s)\}$$

are the generators of cyclic subspaces $V_1, \dots, V_s$ of V, such that $V = V_1 \oplus \cdots \oplus V_s$, and v_j has period $a_j(x)$. One calculates these vectors with the aid of T: if $P^{-1} = (b_{i,j}(x))$, then

$$z_j = \sum_i b_{i,n-s+j}(x) f_i,$$

so

$$v_j = \sum_i b_{i,n-s+j}(T) e_i.$$

Let δ_j denote the degree of $a_j(x)$. Then

$$(v_1, Tv_1, \dots, T^{\delta_1 - 1} v_1; v_2, Tv_2, \dots, T^{\delta_2 - 1} v_2; \dots)$$

is a basis of V with respect to which the matrix of T is in rational canonical form. The reader is asked to fill in some of the details of this discussion in Exercise M.6.3.

Example M.6.10.

The characteristic polynomial and minimal polynomial

Let $A \in \mathrm{Mat}_n(K)$. Write $x - A$ for $xE_n - A$. We define the *characteristic polynomial* of A by $\chi_A(x) = \det(x - A)$. The reader can check that $\chi_A(x)$ is a similarity invariant for A; that is, it is unchanged if A is replaced by a similar matrix. Let V be an n–dimensional vector space over K and let $T \in \mathrm{End}_K(V)$. If A is the matrix of T with respect to some basis of V, define $\chi_T(x) = \chi_A(x)$. It follows from the invariance of χ_A under similarity that χ_T is well–defined (does not depend on the choice of basis) and that χ_T is a similarity invariant for linear transformations. See Exercise M.6.4. $\chi_T(x)$ is called the *characteristic polynomial of* T.

Let A be the matrix of T with respect to some basis of V. Consider the diagonalization of $xE_n - A$ in $\mathrm{Mat}_n(K[x])$,

$$P(xE_n - A)Q = D(x) = \mathrm{diag}(1, 1, \dots, 1, a_1(x), a_2(x), \dots, a_s(x)),$$

where the $a_i(x)$ are the (monic) invariant factors of T. We have

$$\chi_T(x) = \chi_A(x) = \det(xE_n - A) = \det(P^{-1}) \det(D(x)) \det(Q^{-1}).$$

P^{-1} and Q^{-1} are invertible matrices in $\mathrm{Mat}_n(K[x])$, so their determinants are units in $K[x]$, that is nonzero elements of K. Because both $\chi_T(x)$ and $\det(D(x))$ are monic polynomials, it follows that $\det(P^{-1}) \det(Q^{-1}) = 1$, and $\chi_T(x) = \det(D(x)) = \prod_i a_i(x)$. We have proved:

Proposition M.6.11. *The characteristic polynomial of $T \in \mathrm{End}_k(V)$ is the product of the invariant factors of T. The characteristic polynomial of $A \in \mathrm{Mat}_n(K)$ is the product of the invariant factors of A.*

The *minimal polynomial* $\mu_T(x)$ of a linear transformation $T \in \mathrm{End}_K(V)$ is defined to be the largest of the invariant factors of T. Thus $\mu_T(x)$ is the period of the $K[x]$–module determined by T. Since $\mu_T(x)$ is the monic generator of the annihilator of the $K[x]$–module V, it is characterized as the monic polynomial of least degree in $\mathrm{ann}(V)$, that is, the monic polynomial of least degree such that $\mu_T(T) = 0$.

The *minimal polynomial* $\mu_A(x)$ of a matrix $A \in \mathrm{Mat}_n(K)$ is defined to be the largest invariant factor of A. The polynomial $\mu_A(x)$ is characterized as the monic polynomial of least degree such that $\mu_A(A) = 0$.

The following result is a corollary of Proposition M.6.11.

Corollary M.6.12. *(Cayley-Hamilton Theorem) Let $T \in \mathrm{End}_K(V)$*

(a) *The minimal polynomial of T divides the characteristic polynomial of T*

(b) *The minimal polynomial of T has degree at most* $\dim(V)$.

(c) $\chi_T(T) = 0$.

Proof. This is immediate, since $\mu_T(x)$ is the largest invariant factor of T, and $\chi_T(x)$ is the product of all of the invariant factors. ■

Let us make a few more remarks about the relation between the minimal polynomial and the characteristic polynomial. All of the invariant factors of T divide the minimal polynomial $\mu_T(x)$, and $\chi_T(x)$ is the product of all the invariant factors. It follows that $\chi_T(x)$ and $\mu_T(x)$ have the same irreducible factors, but with possibly different multiplicities. Since $\lambda \in K$ is a root of a polynomial exactly when $x - \lambda$ is an irreducible factor, we also have that $\chi_T(x)$ and $\mu_T(x)$ have the same roots, but with possibly different multiplicities. Finally, the characteristic polynomial and the minimal polynomial coincide precisely if V is a cyclic $K[x]$–module; i.e., the rational canonical form of T has only one block.

Of course, statements analogous to Corollary M.6.12, and of these remarks, hold for a matrix $A \in \mathrm{Mat}_n(K)$ in place of the linear transformation T.

The roots of the characteristic polynomial (or of the minimal polynomial) of $T \in \mathrm{End}_K(V)$ have an important characterization. We say that an *nonzero* vector $v \in V$ is an *eigenvector* of T with *eigenvalue* [1] λ, if $Tv = \lambda v$.

Proposition M.6.13. *Let $T \in \mathrm{End}_K(V)$. An element $\lambda \in K$ is a root of $\chi_T(x)$ if, and only if, T has an eigenvector in V with eigenvalue λ.*

Proof. Exercise M.6.6 ■

Exercises M.6

M.6.1. Let $h(x) \in K[x]$ be a polynomial of one variable. Show that there is a polynomial $g(x, y) \in K[x, y]$ such that $h(x) - h(y) = (x - y)g(x, y)$.

M.6.2. Set $w_j = (x-T)f_j = xf_j - \sum_i a_{i,j} f_i$. Show that $\{w_1, \dots, w_n\}$ is linearly independent over $K[x]$.

M.6.3. Verify the following assetions made in the text regarding the computation of the rational canonical form of T. Suppose that F is a free $K[x]$ module, $\Phi : F \longrightarrow V$ is a surjective $K[x]$–module homomorphism, $(y_1, \dots, y_{n-s}, z_1, \dots, z_s)$ is a basis of F, and

$$(y_1, \dots, y_{n-s}, a_1(x)z_1, \dots, a_s(x)z_s)$$

is a basis of $\ker(\Phi)$. Set $v_j = \Phi(z_j)$ for $1 \le j \le s$, and

$$V_j = K[x]v_j = \mathrm{span}(\{p(T)v_j : p(x) \in K[x]\}).$$

(a) Show that $V = V_1 \oplus \cdots \oplus V_s$.

(b) Let δ_j be the degree of $a_j(x)$. Show that $(v_j, Tv_j, \dots, T^{\delta_j - 1}v_j)$ is a basis of V_j. and that the matrix of $T_{|V_j}$ with respect to this basis is the companion matrix of $a_j(x)$.

M.6.4. Show that χ_A is a similarity invariant of matrices. Conclude that for $T \in \mathrm{End}_K(V)$, χ_T is well defined, and is a similarity invariant for linear transformations.

M.6.5. Since $\chi_A(x)$ is a similarity invariant, so are all of its coefficients. Show that the coefficient of x^{n-1} is the negative of the *trace* $\mathrm{tr}(A)$, namely the sum of the matrix entries on the main diagonal of A. Conclude that the trace is a similarity invariant.

[1] These are half-translated German words. The German Eigenvektor and Eigenwert mean "characteristic vector" and "characteristic value."

M.6.6. Show that λ is a root of $\chi_T(x)$ if, and only if, T has an eigenvector in V with eigenvalue λ. Show that v is an eigenvector of T for some eigenvalue if, and only if, the one dimensional subspace $Kv \subseteq V$ is invariant under T.

The next four exercises give an alternative proof of the Cayley-Hamilton theorem. Let $T \in \mathrm{End}_K(V)$, where V is n–dimensional. Assume that the field K contains all roots of $\chi_T(x)$; that is, $\chi_T(x)$ factors into linear factors in $K[x]$.

M.6.7. Let $V_0 \subseteq V$ be any invariant subspace for T. Show that there is a linear operator $\overline{T}$ on V/V_0 defined by

$$\overline{T}(v + V_0) = T(v) + V_0$$

for all $v \in V$. Suppose that $(v_1, \dots, v_k)$ is an ordered basis of V_0, and that

$$(v_{k+1} + V_0, \dots, v_n + V_0)$$

is an ordered basis of V/V_0. Suppose, moreover, that the matrix of $T_{|V_0}$ with respect to $(v_1, \dots, v_k)$ is A_1 and the matrix of $\overline{T}$ with respect to $(v_{k+1}+V_0, \dots, v_n+V_0)$ is A_2. Show that $(v_1, \dots, v_k, v_{k+1}, \dots, v_n)$ is an orderd basis of V and that the matrix of T with respect to this basis has the form

$$\begin{bmatrix} A_1 & B \\ 0 & A_2 \end{bmatrix},$$

where B is some k–by–$(n-k)$ matrix.

M.6.8. Use the previous two exercises, and induction on n to conclude that V has some basis with respect to which the matrix of T is *upper triangular*; that means that all the entries below the main diagonal of the matrix are zero.

M.6.9. Suppose that A' is the upper triangular matrix of T with respect to some basis of V. Denote the diagonal entries of A' by $(\lambda_1, \dots, \lambda_n)$; this sequence may have repetitions. Show that $\chi_T(x) = \prod_i (x - \lambda_i)$.

M.6.10. Let $(v_1, \dots, v_n)$ be a basis of V with respect to which the matrix A' of T is upper triangular, with diagonal entries $(\lambda_1, \dots, \lambda_n)$. Let $V_0 = \{0\}$ and $V_k = \mathrm{span}(\{v_1, \dots, v_k\})$ for $1 \le k \le n$. Show that $T - \lambda_k$ maps V_k into V_{k-1} for all k, $1 \le k \le n$. Show by induction that $(T - \lambda_k)(T - \lambda_{k+1}) \cdots (T - \lambda_n)$ maps V into V_{k-1} for all k, $1 \le k \le n$. Note in particular that $(T - \lambda_1) \cdots (T - \lambda_n) = 0$. Using the previous exercise, conclude that $\chi_T(T) = 0$, the characteristic polynomial of T, evaluated at T, gives the zero transformation.

Remark M.6.14. The previous four exercises show that $\chi_T(T) = 0$, under the assumption that all roots of the characteristic polynomial lie in K. This restriction can be removed, as follows. First, the assertion $\chi_T(T) = 0$ for $T \in \mathrm{End}_K(V)$ is equivalent to the assertion that $\chi_A(A) = 0$ for $A \in \mathrm{Mat}_n(K)$. Let K be any field, and let $A \in \mathrm{Mat}_n(K)$. If F is any field with $F \supseteq K$ then A can be considered as an element of $\mathrm{Mat}_n(F)$. The characteristic polynomial of A is the same whether A is regarded as a matrix with entries in K or as a matrix with entries in F. Moreover, $\chi_A(A)$ is the same matrix, whether A is regarded as a matrix with entries in K or as a matrix with entries in F.

As is explained in Section 8.2, there exists a field $F \supseteq K$ such that all roots of $\chi_A(x)$ lie in F. It follows that $\chi_A(A) = 0$.