Proof: Let B be the rational canonical form of A. By the previous lemma the block diagonal form of B shows that the characteristic polynomial of B is the product of the characteristic polynomials of the companion matrices of the invariant factors of A. By the first part of the lemma above, the characteristic polynomial of the companion matrix $\mathcal{C}_{a(x)}$ for $a(x)$ is just $a(x)$, which implies that the characteristic polynomial for B is the product of the invariant factors of A. Since A and B are similar, they have the same characteristic polynomial, which proves (1). Assertion (2) is immediate from (1) since the minimal polynomial for A is the largest invariant factor of A. The fact that all the invariant factors divide the largest one immediately implies (3). The final assertion is clear from the dictionary between linear transformations of vector spaces and matrices.

Note that part (2) of the proposition is the assertion that the matrix A satisfies its own characteristic polynomial, i.e., $c_A(A) = 0$ as matrices, which is the usual formulation for the Cayley–Hamilton Theorem. Note also that it implies the degree of the minimal polynomial for A has degree at most n, a result mentioned before.

The relations in Proposition 20 are frequently quite useful in the determination of the invariant factors for a matrix A, particularly for matrices of small degree (cf. Exercises 3 and 4 and the examples). The following result (which relies on Exercises 16 to 19 in the previous section and whose proof we outline in the exercises) computes the invariant factors in general.

Let A be an $n \times n$ matrix over the field F. Then $xI - A$ is an $n \times n$ matrix with entries in $F[x]$. The three operations

(a) interchanging two rows or columns

(b) adding a multiple (in $F[x]$) of one row or column to another

(c) multiplying any row or column by a unit in $F[x]$, i.e., by a nonzero element in F,

are called *elementary row and column operations.*

Theorem 21. Let A be an $n \times n$ matrix over the field F. Using the three elementary row and column operations above, the $n \times n$ matrix $xI - A$ with entries from $F[x]$ can be put into the diagonal form (called the *Smith Normal Form* for A)

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & a_1(x) & & & \\ & & & & a_2(x) & & \\ & & & & & \ddots & \\ & & & & & & a_m(x) \end{pmatrix}$$

with monic nonzero elements $a_1(x), a_2(x), \dots, a_m(x)$ of $F[x]$ with degrees at least one and satisfying $a_1(x) \mid a_2(x) \mid \cdots \mid a_m(x)$. The elements $a_1(x), \dots, a_m(x)$ are the invariant factors of A.

Proof: cf. the exercises.

Invariant Factor Decomposition Algorithm: Converting to Rational Canonical Form

As mentioned in the exercises near the end of the previous section, keeping track of the operations necessary to diagonalize $xI - A$ will explicitly give a matrix P such that $P^{-1}AP$ is in rational canonical form. Equivalently, if V is a given $F[x]$-module with vector space basis $[e_1, e_2, \dots, e_n]$, then P defines the change of basis giving the Invariant Factor Decomposition of V into a direct sum of cyclic $F[x]$-modules. In particular, if A is the matrix of the linear transformation T of the $F[x]$-module V defined by x (i.e., $T(e_j) = xe_j = \sum_{i=1}^{n} a_{ij}e_i$ where $A = (a_{ij})$), then the matrix P defines the change of basis for V with respect to which the matrix for T is in rational canonical form.

We first describe the algorithm in the general context of determining the Invariant Factor Decomposition of a given $F[x]$-module V with vector space basis $[e_1, e_2, \dots, e_n]$ (the proof is outlined in the exercises). We then describe the algorithm to convert a given $n \times n$ matrix A to rational canonical form (in which reference to an underlying vector space and associated linear transformation are suppressed).

Explicit numerical examples of this algorithm are given in Examples 2 and 3 following.

Invariant Factor Decomposition Algorithm

Let V be an $F[x]$-module with vector space basis $[e_1, e_2, \dots, e_n]$ (so in particular these elements are generators for V as an $F[x]$-module). Let T be the linear transformation of V to itself defined by x and let A be the $n \times n$ matrix associated to T and this choice of basis for V, i.e.,

$$T(e_j) = xe_j = \sum_{i=1}^{n} a_{ij}e_i \qquad \text{where} \qquad A = (a_{ij}).$$

(1) Use the following three elementary row and column operations to diagonalize the matrix $xI - A$ over $F[x]$, keeping track of the *row* operations used:

  (a) interchange two rows or columns (which will be denoted by $R_i \leftrightarrow R_j$ for the interchange of the i^{th} and j^{th} rows and similarly by $C_i \leftrightarrow C_j$ for columns),

  (b) add a multiple (in $F[x]$) of one row or column to another (which will be denoted by $R_i + p(x)R_j \mapsto R_i$ if $p(x)$ times the j^{th} row is added to the i^{th} row, and similarly by $C_i + p(x)C_j \mapsto C_i$ for columns),

  (c) multiply any row or column by a unit in $F[x]$, i.e., by a nonzero element in F (which will be denoted by uR_i if the i^{th} row is multiplied by $u \in F^{\times}$, and similarly by uC_i for columns).

(2) Beginning with the $F[x]$-module generators $[e_1, e_2, \dots, e_n]$, for each row operation used in (1), change the set of generators by the following rules:

  (a) If the i^{th} row is interchanged with the j^{th} row then interchange the i^{th} and j^{th} generators.

  (b) If $p(x)$ times the j^{th} row is added to the i^{th} row then subtract $p(x)$ times the i^{th} generator from the j^{th} generator (note the indices).

(c) If the i^{th} row is multiplied by the unit $u \in F$ then divide the i^{th} generator by u.

(3) When $xI - A$ has been diagonalized to the form in Theorem 21 the generators $[e_1, e_2, \dots, e_n]$ for V will be in the form of $F[x]$-linear combinations of $e_1, e_2, \dots, e_n$. Use $xe_j = T(e_j) = \sum_{i=1}^n a_{ij}e_i$ to write these elements as F-linear combinations of $e_1, e_2, \dots, e_n$. When $xI - A$ has been diagonalized, the first $n - m$ of these linear combinations are 0 (providing a useful numerical check on the computations) and the remaining m linear combinations are nonzero, i.e., the generators for V are in the form $[0, \dots, 0, f_1, \dots, f_m]$ corresponding precisely to the diagonal elements in Theorem 21. The elements $f_1, \dots, f_m$ are a set of $F[x]$-module generators for the cyclic factors in the invariant factor decomposition of V (with annihilators $(a_1(x)), \dots, (a_m(x))$, respectively):

$$V = F[x]\, f_1 \oplus F[x]\, f_2 \oplus \dots \oplus F[x]\, f_m,$$
$$F[x]\, f_i \cong F[x]/(a_i(x)) \qquad i = 1, 2, \dots, m,$$

giving the Invariant Factor Decomposition of the $F[x]$-module V.

(4) The corresponding *vector space* basis for each cyclic factor of V is then given by the elements $f_i, Tf_i, T^2 f_i, \dots, T^{\deg a_i(x)-1} f_i$.

(5) Write the k^{th} element of the vector space basis computed in (4) in terms of the original vector space basis $[e_1, e_2, \dots, e_n]$ and use the coordinates for the k^{th} column of an $n \times n$ matrix P. Then $P^{-1}AP$ is in rational canonical form (with diagonal blocks the companion matrices for the $a_i(x)$). This is the matrix for the linear transformation T with respect to the vector space basis in (4).

We now describe the algorithm to convert a given $n \times n$ matrix A to rational canonical form, i.e., to determine an $n \times n$ matrix P so that $P^{-1}AP$ is in rational canonical form. This is nothing more than the algorithm above applied to the vector space $V = F^n$ of $n \times 1$ column vectors with standard basis $[e_1, e_2, \dots, e_n]$ (where e_i is the column vector with 1 in the i^{th} position and 0's elsewhere) and T is the linear transformation defined by A and this choice of basis. Explicit reference to this underlying vector space and associated linear transformation are suppressed, so the algorithm is purely matrix theoretic.

Converting an $n \times n$ Matrix to Rational Canonical Form

Let A be an $n \times n$ matrix with entries in the field F.

(1) Use the following three elementary row and column operations to diagonalize the matrix $xI - A$ over $F[x]$, keeping track of the *row* operations used:
   (a) interchange two rows or columns (which will be denoted by $R_i \leftrightarrow R_j$ for the interchange of the i^{th} and j^{th} rows and similarly by $C_i \leftrightarrow C_j$ for columns),
   (b) add a multiple (in $F[x]$) of one row or column to another (which will be denoted by $R_i + p(x)R_j \mapsto R_i$ if $p(x)$ times the j^{th} row is added to the i^{th} row, and similarly by $C_i + p(x)C_j \mapsto C_i$ for columns),
   (c) multiply any row or column by a unit in $F[x]$, i.e., by a nonzero element in F (which will be denoted by uR_i if the i^{th} row is multiplied by $u \in F^\times$, and similarly by uC_i for columns).