A RATIONAL CANONICAL FORM ALGORITHM

K.R. Matthews

Department of Mathematics, University of Queensland
QLD 4072, Australia

January 22, 2002

1 Introduction.

In this note we show how the Jordan canonical form algorithm of Väliaho[8] can be generalized to give the rational canonical form of a square matrix A over an arbitrary field F. If $m_A = p_1^{b_1} \cdots p_t^{b_t}$ is the factorization of the minimum polynomial of A into distinct monic irreducible factors, our objective is to find a non–singular matrix P over F such that

$$P^{-1}AP = H_1 \oplus \cdots \oplus H_t,$$

where

$$H_i = H(p_i^{e_{i1}}) \oplus \cdots H(p_i^{e_{i\gamma_i}})$$

and where the hypercompanion matrix $H(p_i^{e_{ij}})$ is defined by

$$H(p_i^{e_{ij}}) = \begin{bmatrix} C(p_i) & 0 & \cdots & 0 \\ N & C(p_i) & \cdots & 0 \\ 0 & N & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & N & C(p_i) \end{bmatrix}.$$

There are e_{ij} blocks on the diagonal and N is a square matrix of same size as $C(p_i)$, the companion matrix of p_i, where

$$C(p) = \begin{bmatrix} 0 & 0 & & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

if $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$.

Every entry of N is zero, apart from the top right–hand corner, where there is a 1. The overall effect is an unbroken subdiagonal of $1'$s.

In the special case that $p_i = x - \lambda_i$, $H(p_i^{e_{ij}})$ reduces to the elementary Jordan matrix

$$J_{e_{ij}}(\lambda_i) = \begin{bmatrix} \lambda_i & 0 & & & 0 \\ 1 & \lambda_i & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & \lambda_i & 0 \\ 0 & 0 & & 1 & \lambda_i \end{bmatrix}$$

We present our algorithm in terms of linear transformations. However for matrices, the algorithm is easily translated into one which can be used directly by any exact arithmetic matrix calculator which works over F and which computes the minimum polynomial m_A of a square matrix A and factorizes m_A as a product of monic irreducibles over $F[x]$.

Rational canonical forms were first introduced by Frobenius in 1879. (See [3, page 72] for references to this and other early papers.)

Of the many modern proofs of the rational form decomposition, typical are the ones in Friedberg, Insel, Spence [1, pages 339–354], Pearl [5, pages 157–164] and Rotman [7, pages 54–56]. Their proofs are inductive in nature and do not lend themselves to immediate computer implementation.

There is another standard proof based on the Smith canonical form of the matrix $xI - A$, where $A = [T]_\beta$ is the matrix of T relative to a basis β (see Perlis [6, page 162]). However, computationally the resulting algorithm is limited to matrices of small size.

As is well–known (see Jacobson [2, page 188]), V becomes a left–$F[x]$ module if left–$F[x]$ multiplication is defined by

$$fv = f(T)(v), \quad f \in F[x],\ v \in V.$$

With appropriate changes of terminology, our algorithm generalizes to give a proof of the structure theorem for finitely–generated torsion modules M over a principal ideal domain R: simply replace V by M, $F[x]$ by R and replace the scalar multiplication (3), defined below, by the left–module multiplication fv, $f \in R$, $v \in M$.

We finish the paper with an example of an 6×6 matrix over $\mathbb{Z}_3$.

In the interests of brevity, all proofs are omitted and left as exercises. Most are straightforward.

2 Definitions.

Let $T : V \to V$ be a linear transformation over F, with $\dim V = n$. Let $m_T = p_1^{b_1} \cdots p_t^{b_t}$ be the factorization of the minimum polynomial of T into distinct monic irreducible factors. We make crucial use of the vector spaces

$$N_{h,p_i} = \operatorname{Im} p_i^{h-1}(T) \cap \operatorname{Ker} p_i(T), \quad 1 \le i \le t,\ 1 \le h \le b_i, \tag{1}$$

following Väliaho, who dealt with the special case $p_i = x - \lambda_i$.

In particular $N_{1,\,p_i} = \operatorname{Ker} p_i(T)$. Then we have the sequence of subspace containments:

$$N_{1,p_i} \supseteq \cdots \supseteq N_{b_i,p_i} \neq \{0\}. \tag{2}$$

The following result is important for computing a basis for $N_{h,\,p_i}$:

If $\operatorname{Ker} p_i^h(T) = \langle u_1, \ldots, u_r \rangle$, the subspace generated by $u_1, \ldots, u_r$, then

$$N_{h,\,p_i} = \langle p_i^{h-1}(T)(u_1), \ldots, p_i^{h-1}(T)(u_r) \rangle.$$

Let $F_{p_i} = F[x]/(p_i)$ be the field of residue classes mod p_i. Then in addition to being an F–vector space, $N_{h,\,p_i}$ is also an F_{p_i}–vector space if F_{p_i}–scalar multiplication is defined as follows:

Let $\overline{f} = f + (p_i),\ f \in F[x]$ and $v \in N_{h,\,p_i}$. Then

$$\overline{f} v = f(T)(v). \tag{3}$$

Some relevant properties of this scalar multiplication are:

(i) $\overline{f} = \overline{g} \Leftrightarrow p_i \mid f - g$ (that is p_i divides $f - g$).

(ii) Let $n_i = \deg p_i$. Then

$$v = \overline{f_1} w_1 + \cdots + \overline{f_r} w_r \Leftrightarrow v = \sum_{j=1}^{r} \sum_{k=0}^{n_i - 1} c_{jk} T^k(w_j), \quad c_{jk} \in F.$$

(iii) Vectors $w_1, \ldots, w_r$ in $N_{h,\,p_i}$ are F_{p_i}–linearly independent if and only if

$$f_1(T)(w_1) + \cdots + f_r(T)(w_r) = 0 \Rightarrow p_i \mid f_1, \ldots, p_i \mid f_r.$$

This last implication is in turn equivalent to the statement that

$$w_1,\ T(w_1), \ldots, T^{n_i-1}(w_1), \ldots, w_r,\ T(w_r), \ldots, T^{n_i-1}(w_r) \tag{4}$$

are F–linearly independent.

We refer to the expanded array (4) as the *padded* array. It is the means whereby in numerical examples, F_{p_i}–basis calculations can be reduced to F–basis calculations.

From property (iii) we have

$$\nu_{h,\,p_i} = \dim_{F_{p_i}} N_{h,\,p_i} = \frac{1}{\deg p_i} \dim_F N_{h,\,p_i} = \frac{\nu(p_i^h(T)) - \nu(p_i^{h-1}(T))}{\deg p_i}, \tag{5}$$

where $\nu(p_i^h(T))$ denotes the F–nullity of $p_i^h(T)$. (See Mirsky [4, page 161].)

The integers $\nu_{h,\,p_i},\ 1 \le i \le t,\ 1 \le h \le b_i$, form a sequence called the Weyr characteristic (see MacDuffee [3, page 74]). In view of the sequence of containments (2), we have for $1 \le i \le t$, the decreasing sequence of positive integers:

$$\nu_{1,\,p_i} \ge \cdots \ge \nu_{b_i,\,p_i},$$

where $\nu_{1,\,p_i} = \dim_{F_{p_i}} \operatorname{Ker} p_i(T) = \dfrac{\nu(p_i(T))}{\deg p_i}$.

Telescopic cancellation using (5) gives

$$\nu_{1,\,p_i} + \cdots + \nu_{b_i,\,p_i} = \frac{\nu(p_i^{b_i}(T))}{\deg p_i}.$$

We mention that this sum in fact equals a_i, where $p_i^{a_i}$ is the exact power of p_i which divides the characteristic polynomial ch_T. (This emerges as a consequence of taking characteristic polynomials of both sides of (10) in Section 3.)

It is helpful to visualize the above sum as a dot diagram formed by a tower of left–justified rows of dots, where the h–th row from the bottom contains $\nu_{h,\,p_i}$ dots. The height of the tower is b_i, while the width at the bottom is $\gamma_i = \nu_{1,\,p_i}$. For example with $b_i = 4$ and $\nu_{1,\,p_i} = \nu_{2,\,p_i} = 3,\ \nu_{3,\,p_i} = \nu_{4,\,p_i} = 2$, we have the dot diagram

$$\begin{array}{|c|c|c}
\hline
\cdot & \cdot & \multicolumn{1}{c}{} \\
\cline{1-2}
\cdot & \cdot & \multicolumn{1}{c}{} \\
\hline
\cdot & \cdot & \multicolumn{1}{c|}{\cdot} \\
\hline
\cdot & \cdot & \multicolumn{1}{c|}{\cdot} \\
\hline
\end{array}
\qquad
\begin{array}{l}
\nu_{4,\,p_i} \\
\nu_{3,\,p_i} \\
\nu_{2,\,p_i} \\
\nu_{1,\,p_i}
\end{array}$$

The integers represented by the respective columns of dots from left to right, form a decreasing sequence

$$b_i = e_{i1} \ge \cdots \ge e_{i\gamma_i}.$$

These sequences for $1 \le i \le t$ form the Segre characteristic of T (see MacDuffee [3, page 74]). For example, in the above dot diagram, the conjugate partition is $e_{i1} = 4,\ e_{i2} = 4,\ e_{i3} = 2$.

The polynomials $p_i^{e_{ij}},\ 1 \le i \le t,\ 1 \le j \le \gamma_i$ are called the elementary divisors of T.

3 Decomposition of V into indecomposable T–cyclic subspaces.

(Good references for this section are Friedberg, Insel, Spence [1, pages 280–300] and Pearl [5, pages 137–164].)

If $v \in V$, the T–invariant subspace $C_{T,\,v}$ of V defined by

$$C_{T,\,v} = \{f(T)(v) | f \in F[x]\}$$

is called the T–cyclic subspace generated by v. The minimum polynomial $m_{T,\,v}$ of v is the monic polynomial f of least degree such that $f(T)(v) = 0$. If $v \ne 0$, then $m = \deg m_{T,\,v} > 0$ and $C_{T,\,v}$ has a basis β:

$$v,\ T(v), \ldots, T^{m-1}(v)$$

called a T–cyclic basis. If $W = C_{T,v}$ and T_W denotes the restriction of T to W, then $[T_W]_\beta = C(m_{T,v})$.

In the special case where $m_{T,v} = p^e$, where p is a monic irreducible polynomial of degree n, $C_{T,v}$ has another basis β':

$$\begin{array}{cccc} v, & T(v), & \ldots, & T^{n-1}(v) \\ p(T)(v), & Tp(T)(v), & \ldots, & T^{n-1}p(T)(v) \\ \vdots & \vdots & \vdots & \vdots \\ p^{e-1}(T)(v), & Tp^{e-1}(T)(v), & \ldots, & T^{n-1}p^{e-1}(T)(v) \end{array}$$

called a canonical basis. Here $[T_W]_\beta = H(p^e)$.

The well–known primary decomposition theorem (see Friedberg, Insel, Spence [1, pages 342–343]) states that

$$V = \operatorname{Ker} p_1^{b_1}(T) \oplus \cdots \oplus \operatorname{Ker} p_t^{b_t}(T). \tag{6}$$

We will give an algorithm which decomposes each $\operatorname{Ker} p_i^{b_i}(T)$ into a direct sum of indecomposable T–cyclic subspaces:

$$\operatorname{Ker} p_i^{b_i}(T) = \bigoplus_{j=1}^{\gamma_i} C_{T,v_{ij}}, \quad \text{where} \quad m_{T,v_{ij}} = p_i^{e_{ij}}. \tag{7}$$

Consequently in view of (6), we have a decomposition of V as a direct sum of indecomposable T–cyclic subspaces:

$$V = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} C_{T,v_{ij}}. \tag{8}$$

Then if β_{ij} is the canonical basis for $C_{T,v_{ij}}$ and

$$\beta = \bigcup_{i=1}^{t} \bigcup_{j=1}^{\gamma_i} \beta_{ij}, \tag{9}$$

then β is a basis for V with the property that

$$[T]_\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} H(p_i^{e_{ij}}). \tag{10}$$

We can now apply the result to the special case $T = T_A : V_n(F) \to V_n(F)$, where $V_n(F)$ is the F–space of n–dimensional column vectors over F, $A \in M_{n\times n}(F)$ and $T_A(X) = AX$. If P is the non–singular matrix whose columns are the respective members of the basis β defined in (9):

$$P = [v_{11}|\cdots|v_{1\gamma_1}|\cdots|v_{t1}|\cdots|v_{t\gamma_t}],$$

then

$$P^{-1}AP = [T_A]_\beta = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{\gamma_i} H(p_i^{e_{ij}}).$$

4 Constructing the vectors v_{ij}.

The motivation for the construction of the vectors v_{ij} comes from a uniqueness result for the elementary divisors, which involves the F_{p_i}–vector spaces $N_{h,\,p_i}$. For we see that in any decomposition (7), we must have

$$m_T = p_1^{e_{i1}} \cdots p_t^{e_{i\gamma_i}},$$

thereby determining the polynomials $p_1, \ldots, p_t$ as the distinct monic irreducible factors of m_T. Also for each $i,\ 1 \le i \le t$, if $1 \le h \le b_i$, it is easy to prove that $N_{h,\,p_i}$ has the F_{p_i}–basis

$$p_i^{e_{i1}-1} v_{i1}, \ldots, p_i^{e_{ij_h}-1} v_{ij_h}, \tag{11}$$

where $e_{i1}, \ldots, e_{ij_h}$ are the integers in the sequence $e_{i1}, \ldots, e_{i\gamma_i}$ which are not less than h.

There are consequently $\dim_{F_{p_i}} N_{h,\,p_i} = \nu_{h,\,p_i}$ such integers and hence the number of integers $e_{i1}, \ldots, e_{i\gamma_i}$ equal to h is equal to $\nu_{h,\,p_i} - \nu_{h+1,\,p_i}$, which depends only on T. In other words, for each i, the sequence $e_{i1}, \ldots, e_{i\gamma_i}$ depends only on T.

In particular, $\operatorname{Ker} p_i(T)$ possesses a special type of F_{p_i}–basis

$$p_i^{e_{i1}-1}(T)(v_{i1}), \ldots, p_i^{e_{i\gamma_i}-1}(T)(v_{i\gamma_i}) \tag{12}$$

with the property that the vectors (11) with $j_h = \nu_{h,\,p_i}$, form an F_{p_i}–basis for $N_{b_i,\,p_i},\ 1 \le h \le b_i$.

In fact such a basis is easy to construct. We start with a F_{p_i}–basis for $N_{b_i,\,p_i}$, extending it to bases for the successive distinct subspaces in the sequence

$$N_{b_i,\,p_i} \subseteq \cdots \subseteq N_{1,\,p_i},$$

until we eventually reach an F_{p_i}–basis for $\operatorname{Ker} p_i(T)$ of the required form (12).

It is then straightforward to prove that the secondary decomposition (7) follows as a consequence. (The reader is urged to verify this statement in the particular case of the earlier dot diagram. A proof by induction of the general case, should then suggest itself.)

We illustrate the construction of the F_{p_i}–basis (12) using the earlier dot diagram: here $e_{i1} = 4,\ e_{i2} = 4,\ e_{i3} = 2$.

First choose an F_{p_i}–basis $p_i^3(T)(v_{i1}),\ p_i^3(T)(v_{i2})$ for $N_{4,\,p_i} = N_{3,\,p_i}$. Then extend this to an F_{p_i}–basis $p_i^3(T)(v_{i1}),\ p_i^3(T)(v_{i2}),\ p_i(T)(v_{i3})$ for $N_{2,\,p_i} = N_{1,\,p_i}$. Then $\operatorname{Ker} p_i^4(T) = C_{T,\,v_{i1}} \oplus C_{T,\,v_{i2}} \oplus C_{T,\,v_{i3}}$, where $m_{T,\,v_{i1}} = p_i^4 = m_{T,\,v_{i2}}$ and $m_{T,\,v_{i3}} = p_i^2$.

5 A numerical example.

Let $A \in M_{6\times 6}(\mathbb{Z}_3)$:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 2 & 2 \\ 2 & 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \in M_{6\times 6}(\mathbb{Z}_3).$$

Here $m_A = p_1^2,\ p_1 = x^2 + x + 2 \in F[x],\ F = \mathbb{Z}_3,\ p_1(A) = A^2 + A + 2I_6$.

$$p_1(A) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 1 & 0 \\ 0 & 1 & 2 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \nu(p_1(A)) = 4,\ \nu_{1,\,p_1} = \tfrac{\nu(p_1(A))}{\deg p_1} = 2.$$

$$p_1^2(A) = 0,\ \nu(p_1^2(A)) = 6,\ \nu_{2,\,p_1} = \tfrac{\nu(p_1^2(A)) - \nu(p_1(A))}{\deg p_1} = \tfrac{6-4}{2} = 1.$$

Hence we have a corresponding F_{p_1}–dot diagram:

$$\begin{array}{ll} \boxed{\cdot} & \nu_{2,\,p_1} \\ \boxed{\cdot}\boxed{\cdot} & \nu_{1,\,p_1} \end{array}$$

We have to find an F_{p_1}–basis $p(A)v_{11}$ for $N_{2,\,p_1}$ and extend this to an F_{p_1}–basis $p_1(A)v_{11},\ v_{12}$ for $N(p_1(A)) = \operatorname{Ker} p_1(T_A)$.

An F–basis for $N(p_1^2(A))$ is $E_1, \ldots, E_6$, the standard basis for $V_6(F)$. Then

$$N_{2,\,p_1} = \langle p_1(A)E_1, \ldots, p_1(A)E_6\rangle = \langle p_1(A)E_2\rangle.$$

Thus $p_1(A)E_2$ is an F_{p_1}–basis for $N_{2,\,p_1}$ so we can take $v_{11} = E_2$.

We find the columns of the following matrix form an F–basis for $N(p_1(A))$:

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

We place $p_1(A)E_2$ in front of this matrix and then pad the resulting matrix to get

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

The first four columns $p_1(A)E_2$, $Ap_1(A)E_2$, E_1, AE_1 of this matrix form an F–basis for $N(p_1(A))$ and hence $p_1(A)E_2$, E_1 form an F_{p_1}–basis for $N(p_1(A))$. So we can take $v_{12} = E_1$.

Then $V_6(\mathbb{Z}_3) = N(p_1^2(A)) = C_{T_A, v_{11}} \oplus C_{T_A, v_{12}}$ and joining canonical bases v_{11}, Av_{11}, $p_1(A)v_{11}$, $Ap_1(A)v_{11}$ for $C_{T_A, v_{11}}$ and v_{12}, Av_{12} for $C_{T_A, v_{12}}$, gives a basis v_{11}, Av_{11}, $p_1(A)v_{11}$, $Ap_1(A)v_{11}, v_{12}$, Av_{12} for $V_6(\mathbb{Z}_3)$.

Finally, if P is the non–singular matrix whose columns are the respective members of this basis, we can transform A into a direct sum of hypercompanion matrices:

$$P^{-1}AP = H(p^2) \oplus H(p) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix},$$

where

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

References

[1] S.H. Friedberg, A.J. Insel, L.E. Spence, *Linear Algebra*, Prentice–Hall, New Jersey, 1979.

[2] N. Jacobson, *Basic algebra I*, W.H. Freeman and Company, 1974.

[3] C.C. MacDuffee, *The theory of matrices*, Chelsea Publishing Company, New York, 1946.

[4] L. Mirsky, *An introduction to linear algebra*, Oxford University Press, London, 1961.

[5] M. Pearl, *Matrix theory and finite mathematics*, McGraw–Hill, New York, 1973.

[6] S. Perlis, *Theory of matrices*, Addison–Wesley, Second edition, Reading, Massachusetts, 1958.

[7] J. Rotman, *The theory of groups*, Allyn and Bacon, Second edition, Boston, 1973.

[8] H. Väliaho, *An elementary approach to the Jordan canonical form of a matrix*, Amer. Math. Monthly 93 (1986), 711–714.