# Drazin Inverses and Canonical Forms in $M_n(Z/h)$

Robert E. Hartwig
*Mathematics Department*
*North Carolina State University*
*Raleigh, North Carolina 27650*

Submitted by Hans Schneider

ABSTRACT

The existence and construction of the Drazin inverse of a square matrix over the ring $Z/h$ is considered. The canonical forms of matrices over this ring are used to facilitate the computation of this type of generalized inverse.

## 1. INTRODUCTION

In this note we wish to investigate the existence and construction of the Drazin inverse of a square matrix over the commutative ring $Z/h$, that is, the integers modulo $h$. The motivation for this investigation comes from the theory of cryptography, into which the concept of a Drazin inverse has recently been introduced [8, 12].

The Drazin inverse is the unique solution, if any, to the matrix equations

$$A^kXA=A^k,\qquad XAX=X,\qquad AX=XA,\tag{1.1}$$

for some $k\geqslant 0$. The smallest value of $k$ for which a solution exists is called the *index* of $A$. If $\operatorname{index}(A)\leqslant 1$, $A$ is said to have a *group inverse*, which becomes the inverse of $A$ in case $\operatorname{index}(A)=0$. We shall assume familiarity with the basic definitions and theory of this type of generalized inverse, as given in [3], [1], [2]. As usual, we shall denote the Drazin inverse of $A$ by $A^d$, and the group inverse, if any, by $A^{\#}$. Moreover we shall use $M_n(\mathfrak{R})$ or $\mathfrak{R}_{n\times n}$ to denote the ring of $n\times n$ matrices over a ring $\mathfrak{R}$. A ring for which every element has a Drazin inverse is usually called a strong $\pi$-regular ring (s$\pi$r ring for short). A matrix $A_{m\times n}$ is called *regular* if there exists a matrix $A^-$ such that $AA^-A=A$. We shall employ the standard notation of $R(\cdot)$, $\mathrm{RS}(\cdot)$, $N(\cdot)$, and $\rho(\cdot)$, to denote the range, rowspace, nulspace, and rank (in the

© Elsevier North Holland, Inc., 1981
52 Vanderbilt Ave., New York, NY 10017 0024-3795/81/030205+29$02.50

sense of McCoy [13]), for the matrix $(\cdot)$. Similarity will be indicated by $\approx$, and isomorphisms by $\cong$, while equivalence and row equivalence will be denoted by $\sim$ and $\underset{\text{row}}{\sim}$ respectively.

On factoring the integer $h$ into its primary factors, $h=p_1^{m_1}p_2^{m_2}\cdots p_s^{m_s}$, with $p_i$ distinct primes, it suffices to consider $M_n(Z/p^m)$ for some prime $p$ and some integer $m\geqslant 1$. Indeed, it follows that if $A\equiv A_k \pmod{p^{m_k}}$, $k=1,\dots,s$, and $A=A_1\oplus A_2\oplus\cdots\oplus A_s$, then $A^d=A_1^d\oplus A_2^d\oplus\cdots\oplus A_s^d$, from which $A^d$ may be found using the Chinese remainder theorem [11]. Moreover, it is easily seen that $\rho(A)=\min\rho(A_i)$ over $Z/p_i^{m_i}$, $i=1,2,\dots,s$. Hence from now on we shall concentrate on $M_n(\mathfrak{R})$ where $\mathfrak{R}=Z/p^m$.

Before we embark on the search for $A^d$, let us recapitulate some of the salient features about $\mathfrak{R}=Z/p^m$ and $M_n(\mathfrak{R})$.

First, every element $a\in R$ either is a unit or is divisible by $p$, and may *uniquely* be written modulo $p^m$ as polynomial in $p$:

$$a=\alpha_0+\alpha_1p+\alpha_2p^2+\cdots+\alpha_{m-1}p^{m-1} \tag{1.2}$$

where $0\leqslant\alpha_i<p$ for all $i=0,\dots,m-1$. If $\alpha_0=\alpha_1=\cdots=\alpha_{t-1}=0\neq\alpha_t$ and $\alpha_k\neq 0=\alpha_{k+1}=\cdots=\alpha_{n-1}$, then $k$ is called the *degree* $\partial a$ of $a$, while $t$ is called the *order* $\delta(a)$ of $a$. Clearly, if $\delta(a)=t$, then $a=up^t$ for some unit $u$, and conversely. Also, $a$ is a unit if and only if $\delta(a)=0$, while $a=0$ exactly when $\delta(a)=m$. Second, $\mathfrak{R}$ is a local ring, with a unique maximal ideal $I_p$, consisting of all multiples of $p$. It further follows that $\mathfrak{R}/I_p\cong Z/p$. Third, the only regular elements in $\mathfrak{R}$ are the units together with zero. Fourth, every matrix in $M_n(Z/p^m)$ may be considered as a polynomial in $p$ with matrix coefficients. That is, $A\in M_n(Z/p^m)$ implies

$$A=A(p)=A_0+A_1p+A_2p^2+\cdots+A_{m-1}p^{m-1}, \tag{1.3}$$

where each $A_i$ is unique, and may be thought of as a matrix over the field $\mathbb{F}=Z/p$. In other words, we may embed $M_n(Z/p^m)$ into $M_n(\mathbb{F}[\lambda])$. From (1.3) a variety of useful facts follow at once. In particular,

(1) $A$ is invertible if and only if $A_0$ is invertible,

(2) $A$ is nilpotent if and only if $A_0$ is nilpotent,

(3) if $A^{\#}$ exists, so does $A_0^{\#}$,

(4) linearly independent columns (rows) in $A$ will correspond to linearly independent columns (rows) in $A_0$,

(5) a $k\times k$ minor in $A$ is a unit if and only if the corresponding $k\times k$ minor in $A_0$ is a unit,

(6) $\rho(A)=\rho(A_0)$, and both equal the size of the largest unit minor in $A$ or $A_0$.

We shall see that to a large extent, the behavior of $A$ is determined by the character of $A_0$, but of course not entirely. In particular for nilpotent $A_0$, the index of nilpotency for $A$ may be as large as $mn$, as seen from the matrix

$$A=\begin{bmatrix} 0 & p \\ 1+p & 0 \end{bmatrix}.$$

We shall further see that the search for $A^d$ will involve much of the general theory of $M_n(Z/p^m)$, and that a variety of techniques may be used in this investigation, such as:

(1) perturbation theory,
(2) $\lambda$-matrix theory,
(3) the theory of block triangular Toeplitz matrices,
(4) the local ring theory.

## 2. EXISTENCE OF $A^d$ AND THE FITTING DECOMPOSITION

The existence of $A^d$ for any $A\in M_n(Z/p^m)$, is easily established, once one recalls that the matrix ring $M_n(\mathfrak{R})$ is *finite*, with $N=p^{mn^2}$ elements. Hence the chains $R(A)\supseteq R(A^2)\supseteq R(A^3)\supseteq\cdots$ and $\mathrm{RS}(A)\supseteq \mathrm{RS}(A^2)\supseteq \mathrm{RS}(A^3)\supseteq\cdots$ become stationary, which suffices for $A^d$ to exist [7]. Indeed, if $A^{k+1}X=A^k$ and $YA^{l+1}=A^l$, then $A^d=A^kX^{k+1}=Y^{l+1}A^1=YA^kX^k=Y^lA^lX$. We may, however, say more. Indeed, in the sequence $\{I, A, A^2,\dots, A^N\}$ there must be two identical matrices, say $A^k=A^{k+r}$ with $0\leqslant k<k+r\leqslant N$ and $1\leqslant r\leqslant N$. Hence $A^k=A^{k+rt}$ for all $t\geqslant 0$, and since $0\leqslant k\leqslant N-1$, we get $A^{N-1}=A^{N-1+rt}$. Now $r\leqslant N$, and thus $r$ divides $N!$. This means that we may take $rt=N!$ and $A^{N-1}=A^{N-1+N!}$. It now follows [7] that $A^d$ exists and is given by

$$A^d=A^{N!(N-1)-1}, \tag{2.1}$$

for *all* $A\in M_n(Z/p^m)$. In other words, this expression gives an *a priori* bound on the exponent needed to compute $A^d$. We note in passing that the above proof and construction work equally well in any finite ring, with a slight modification needed if the ring has no unity element. The expression (2.1) for $A^d$ is clearly only of theoretical interest because the value of $N!(N-1)-1$ is far too large to be of practical use; even if $m=n=p=2$ we get $256!\times 255-1$. We shall therefore endeavor in the remaining sections of this paper to develop more realistic algorithms for computing $A^d$ in $M_n(\mathfrak{R})$.

The foremost theoretical consequence of the existence of $A^d$ is that Fitting's decomposition is valid.

THEOREM 2.1. *Let $\mathfrak{R}$ be a ring with 1. Then the following are equivalent*:

(i) $A \in M_n(\mathfrak{R})$ *implies that*

$$A = Q\begin{bmatrix} U & 0 \\ 0 & \eta \end{bmatrix}Q^{-1}$$

*for some invertible matrices $Q$, $U$ and some nilpotent matrix $\eta$.*

(ii) *$M_n(\mathfrak{R})$ is $s\pi r$, and for every idempotent matrix $E \in M_n(\mathfrak{R})$ the range $R(E)$ has a basis.*

*Proof.* (i)⇒(ii):

$$\text{If } A = Q\begin{bmatrix} U & 0 \\ 0 & \eta \end{bmatrix}Q^{-1},$$

then obviously

$$A^d = Q\begin{bmatrix} U^{-1} & 0 \\ 0 & 0 \end{bmatrix}Q^{-1}.$$

Now if

$$E^2 = E = Q\begin{bmatrix} U & 0 \\ 0 & \eta \end{bmatrix}Q^{-1},$$

then $U = I$ and $\eta = 0$, ensuring that

$$Q\begin{bmatrix} I \\ 0 \end{bmatrix} = Q_1$$

will form a basis matrix for $R(E)$.

(ii)⇒(i): This proof using mappings is identical to the case where $\mathfrak{R}$ is a field. A pure matrix proof is the following. Let $E = AA^d$, and let $Q_1, Q_2$ be basis matrices for $R(E)$ and $R(I-E)$ respectively. Then $Q = [Q_1, Q_2]$ is a basis matrix for $\mathfrak{R}^n$ and hence is invertible. Now since $R(E)$ and $R(I-E)$ are invariant, we have $AQ_1 = Q_1U$, and $AQ_2 = Q_2\eta$ for some square $U$ and $\eta$. Hence if $k = \text{index}(A)$, then $A^k(I - AA^d) = 0$, so that $0 = A^kQ_2 = Q_2\eta^k$, which by the independence of the columns in $Q_2$ forces $\eta^k = 0$. Also

$$\begin{aligned}(A^k + I - AA^d)[Q_1, Q_2] &= [A^kQ_1, Q_2]\\ &= [Q_1, Q_2]\begin{bmatrix} U^k & 0\\ 0 & I\end{bmatrix}.\end{aligned}$$

But $A^k + I - E$ is invertible with inverse $(A^d)^k + I - E$, which means that $U^k$ and hence $U$ is invertible. ■

The condition that the range of every idempotent $E \in M_n(\mathfrak{R})$ has a basis *is* indeed satisfied for the ring $\mathfrak{R} = Z/p^m$. This is basically a consequence of the local ring structure of $Z/p^m$[14, p. 101]. A more elementary and constructive proof is the following.

Lemma 2.2. *Let $\mathfrak{R}$ be a ring with* 1, *such that every nonzero idempotent matrix $E \in M_n(\mathfrak{R})$ has a unit entry. Then*:

(i) $$E \approx \begin{bmatrix} I_r & 0\\ 0 & 0\end{bmatrix}. \tag{2.2}$$

(ii) *$R(E)$, $\mathrm{RS}(E)$, and $N(E)$ have bases.*

(ii) *For a regular matrix $A$, all of $R(A)$, $\mathrm{RS}(A)$, and $N(A)$ have bases.*

*Proof.* (i): Let $E_1^2 = E_1 \neq 0$, and consider $(I - E_1)E_1 = 0$. Since $I - E_1$ is idempotent, it has a unit entry. Hence there is a row $\boldsymbol{\lambda}^T = [\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots, \lambda_n]$, such that $\lambda_k$ is a unit and $\boldsymbol{\lambda}^T E = \mathbf{0}^T$. Now let

$$R = \begin{bmatrix} 1 & & & & & & & & \\ & 1 & & & & & & & \\ & & 1 & & & & & & \\ & & & 1 & & & & & \\ \lambda_1 & \cdot & \cdot & \cdot & \lambda_k & \cdot & \cdot & \cdot & \lambda_n \\ & & & & & 1 & & & \\ & & & & & & 1 & & \\ & 0 & & & & & & 1 & \\ & & & & & & & & 1 \end{bmatrix}.$$

Then $R$ is a unit and

$$RE_1=\begin{bmatrix} & & ? & & \\ 0 & \cdots & 0 & \cdots & 0 \\ & & ? & & \end{bmatrix}k.$$

Consequently

$$R_1E_1=\begin{bmatrix} & ? & \\ 0 & \cdots & 0 \end{bmatrix}$$

for some unit $R_1$. This gives

$$R_1E_1R_1^{-1}=\begin{bmatrix} E_2 & F_2 \\ 0 & 0 \end{bmatrix},$$

where $E_2^2=E_2$ and $F_2=E_2F_2$. Hence

$$\begin{bmatrix} I & F_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} E_2 & F_2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} I & -F_2 \\ 0 & I \end{bmatrix}=\begin{bmatrix} E_2 & 0 \\ 0 & 0 \end{bmatrix}$$

and so

$$E_1\approx\begin{bmatrix} E_2 & 0 \\ 0 & 0 \end{bmatrix}.$$

This algorithm may now be repeated with $E_2$, $E_3$, etc., and will stop only when we get $E_k=I$. Again, since

$$\begin{bmatrix} I & F_k \\ 0 & 0 \end{bmatrix}\approx\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix},$$

the conclusion follows.

(ii): If

$$E[Q_1,Q_2]=[Q_1,Q_2]\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

then $Q_1$ will yield a basis for $R(E)$ and $Q_2$ a basis for $N(E)$. Indeed, the columns of $Q_2$ are clearly independent, and if $E\mathbf{x}=0$, then $EQ\mathbf{y}=0$ with

$$\mathbf{y}=Q^{-1}\mathbf{x}=\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}.$$

Hence

$$Q\begin{bmatrix} I \\ 0 \end{bmatrix}\mathbf{y}=0 \quad\Rightarrow\quad \mathbf{y}_1=0$$

and thus

$$\mathbf{x}=Q\begin{bmatrix} 0 \\ \mathbf{y}_2 \end{bmatrix}=Q_2\mathbf{y}_2.$$

This means that the columns of $Q_2$ also span $N(E)$. Similarly for $R(E)$ and $\mathrm{RS}(E)$.

(iii): Let $AA^-A=A$. Then $AA^-$ and $A^-A$ are idempotents, and $R(A)=R(AA^-)$, $\mathrm{RS}(A)=\mathrm{RS}(A^-A)$, and $N(A)=R(I-A^-A)$. Hence all three modules have bases, by part (ii). ■

The conditions of Lemma 2.2 are easily seen to hold for $Z/p^m$.

Corollary 2.3. *If $0\neq A\in M_n(Z/p^m)$ has a group inverse, then $A$ has a unit entry.*

*Proof.* If not, then $A=pX$, $A^m=0$, and hence $A=0$. In particular, every nonzero idempotent matrix has a unit entry. ■

We may derive several useful results from the Fitting decomposition:

Corollary 2.4. *Let $A=A_0+A_1p+\cdots+A_{m-1}p^{m-1}\in M_n(Z/p^m)$, with $A_i$ over $Z/p$. Let $\Delta_{A_0}$ and $\psi_{A_0}=\lambda^{k_0}(?)$ denote the characteristic and minimal polynomials of $A_0$, with $k_0=\mathrm{index}(A_0)$. Then*

(i) $$\mathrm{index}(A)\leqslant k_0m, \tag{2.3}$$

(ii) $[\psi_{A_0}(A)]^m=[\Delta_{A_0}(A)]^m=0$.

*Proof.* (i): Let

$$A\approx\begin{bmatrix} U & 0 \\ 0 & \eta \end{bmatrix},$$

with $U=U_0+U_1p+\cdots\ \eta=\eta_0+\eta_1p+\cdots$, and with $U_i, \eta_i$ over $Z/p$. Then $k_0=\text{index}(\eta_0)=\text{index}(A_0)$ and $\eta^{k_0}=pY$ for some $Y$. This gives $\eta^{k_0m}=0$, and thus $\text{index}(A)=\text{index}(\eta)\leqslant k_0m$.

(ii):

$$\psi_{A_0}(A)\approx\begin{bmatrix}\psi_{A_0(U)} & 0\\ 0 & \psi_{A_0}(\eta)\end{bmatrix}.$$

Now $\psi_{A_0}(U)=\psi_{A_0}(U_0)+p(?)$ and $\psi_{A_0}(\eta)=\psi_{A_0}(\eta_0)+p(?)$. But $\psi_{A_0}(A_0)=0$, and thus $\psi_{A_0}(U_0)=0$ and $\psi_{A_0}(\eta_0)=0$. Hence $p$ divides $\psi_{A_0}(A)$ or $[\psi_{A_0}(A)]^m=0$. Since $\psi_{A_0}(\lambda)|\Delta_{A_0}(\lambda)|\Delta_A(\lambda)$, the remaining identity follows. The latter equality in (ii) may also be seen from the Cayley-Hamilton theorem and the fact that

$$\Delta_A(\lambda)=\det(\lambda I-A(p))=\det(\lambda I-A_0-pB)$$

$$=\Delta_{A_0}(\lambda)+pf_1(\lambda)+p^2f_2(\lambda)+\cdots \tag{2.4}$$

for some polynomials $f_i(\lambda)$ over $Z/p$. ■

In general the index of $A$ may be as large as $mk_0$, as seen from

$$A=\begin{bmatrix}0 & p\\ 1+p & 0\end{bmatrix},$$

which has index $2m$. Moreover this example shows that the *lowest* power of $A$ which will be regular in general is $mk_0$.

## 3. PRACTICAL METHODS FOR COMPUTING $A^d$

Let us now turn to some practical methods for calculating $A^d$. In general let us write $A=A_0+pB=C_0+N_0+pB$, where $A\equiv A_0 \bmod p^m$, $C_0=A_0^2A_0^d$ is the *core* of $A_0$, and $N_0=A_0(I-A_0A_0^d)$ is the *nilpotent part* of $A_0$. Since we have two types of nilpotent matrices to contend with, our strategy shall be to try and eliminate one of them by powering, and then treat the remaining special case by iteration. The reason for this is that it is difficult to handle $N_0$ and $pB$ simultaneously.

*Method I. The Characteristic-Polynomial Method*

This method is similar to the one used for matrices over a field, with one extra twist added to remove the extra nilpotency present. First compute the characteristic polynomial $\Delta_A(\lambda)=\lambda^n-\sigma_1\lambda^{n-1}+\cdots(-1)^n\sigma_n$, where

$$\sigma_k=\sum\left[\text{all}\binom{n}{k}\ k\times k\text{ principal minors of }A\right].$$

Then expand $\Delta_A(\lambda)=\Delta_{A_0}(\lambda)+p\cdot g(\lambda)$, where $\Delta_{A_0}(\lambda)=\lambda^n+\cdots+a_k\lambda^k$, with $a_k$ *invertible*. Then by (2.3), $[\Delta_{A_0}(\lambda)]^m=0$. We now expand $\Delta_{A_0}(\lambda)^m$ as $a_k^m\lambda^{mk}[1-\lambda q(\lambda)]$, compute $\lambda^{mk}q(\lambda)^{mk+1}$, and reduce it modulo $\Delta_A(\lambda)$. This yields $A^d$ as a polynomial in $A$. In actual practice this method is rather tedious to do by hand, and is more suitable for machine calculations. It should be clear that this method works for any annihilating polynomial for $A_0$ that is available. Generally speaking though, $\Delta_{A_0}$ is the only such polynomial available, and it is obviously no more work to compute $\Delta_A$ than it is to compute $\Delta_{A_0}(\lambda)$.

Before we continue with the general case, several preliminary results will be needed. We shall, in particular, use Roth's removal rule [15] to handle the general twofold nilpotency, and we shall consider the special cases where $A_0^\#$ or $A^\#$ exists.

LEMMA 3.1. *If $A$ is invertible and* $\begin{bmatrix} A & 0 \\ B & D\end{bmatrix}$ *is regular, then $D$ is regular.*

*Proof.* This is easy, and is left as an exercise. ■

COROLLARY 3.2. *If $A$ is invertible and* $\begin{bmatrix} A & 0 \\ B & pD\end{bmatrix}$ *is regular, then $D=0$.*

LEMMA 3.3. *If $A$ is invertible and $D$ is nilpotent, then*

(i) *$AX-XD=C$ has a unique solution*

$$X=\sum_{r=0}^{n-1}A^{-r-1}CD^r, \tag{3.1}$$

(ii) *$YA-DY=B$ has a unique solution*

$$Y=\sum_{r=0}^{n-1}D^rBA^{-r-1}. \tag{3.2}$$

*Proof.* Iterate $X=A^{-1}C+A^{-1}XD$ and $Y=BA^{-1}+DYA^{-1}$. ■

PROPOSITION 3.4. *If $A$ is invertible and $D$ is nilpotent of index $k$, then*

$$\text{(i)}\quad \begin{bmatrix} A & C \\ B & D \end{bmatrix} \approx \begin{bmatrix} A+XB & -XBX \\ B & D-BX \end{bmatrix}, \quad \textit{with}\quad X=\sum_{r=0}^{k-1} A^{-r-1}CD^{-r},$$

$$\text{(ii)}\quad \begin{bmatrix} A & C \\ B & D \end{bmatrix} \approx \begin{bmatrix} A+CY & C \\ -YCY & D-YC \end{bmatrix}, \quad \textit{with}\quad Y=\sum_{r=0}^{k-1} D^{-r}BA^{-r-1}. \tag{3.3}$$

*Proof.* (i): Let $X$ be the solution to $AX-XD=C$, as defined in (3.1). Then form

$$\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}\begin{bmatrix} A & C \\ B & D \end{bmatrix}\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix}.$$

(ii): Form

$$\begin{bmatrix} I & 0 \\ -Y & I \end{bmatrix}\begin{bmatrix} A & C \\ B & D \end{bmatrix}\begin{bmatrix} I & 0 \\ Y & I \end{bmatrix},$$

with $Y$ the unique solution to $YA-DY=B$ as given in (3.2). ■

COROLLARY 3.5. *If $A$ is invertible and $D$ is nilpotent, then*

$$\text{(i)}\quad \begin{bmatrix} A & pC \\ B & D \end{bmatrix} \approx \begin{bmatrix} A' & 0 \\ B & D' \end{bmatrix},$$

$$\text{(ii)}\quad \begin{bmatrix} A & C \\ pB & D \end{bmatrix} \approx \begin{bmatrix} A'' & C \\ 0 & D'' \end{bmatrix}, \tag{3.4}$$

$$\text{(iii)}\quad \begin{bmatrix} A & pC \\ pB & D \end{bmatrix} \approx \begin{bmatrix} A''' & 0 \\ 0 & D''' \end{bmatrix},$$

*where $A'$, $A''$, $A'''$ are invertible and $D'$, $D''$, $D'''$ are nilpotent.*

*Proof.* (i): From (3.3) (i), we see that if $p$ divides $C$, then $p$ divides $X$. Hence

$$\begin{bmatrix} A & pC \\ B & D \end{bmatrix} \approx \begin{bmatrix} A & p^2(?) \\ B & D \end{bmatrix},$$

with $A$ invertible and $D$ nilpotent. Repeating this algorithm a sufficient number of times, we see that

$$\begin{bmatrix} A & pC \\ B & D \end{bmatrix} \approx \begin{bmatrix} A' & 0 \\ B & D' \end{bmatrix},$$

with $A'$ invertible and $D'$ nilpotent.

(ii): This follows analogously from (3.3) (ii).

(iii): Combine parts (i) and (iv). ■

It should be noted that this time $XBX$ and $YCY$ are of order $p^3$, and hence that the iteration converges to zero much *faster*. Moreover, part (iii) yields a Fitting decomposition of the matrix $\begin{bmatrix} A & pC \\ pB & D \end{bmatrix}$, from which the Drazin inverse may be computed.

It is important to observe that if $D$ is *also* of order $p$, then Roth's removal rule is *not* needed. We may then simply use the *first* terms in the expansions (3.1) and (3.2), namely $X = A^{-1}C$ and $Y = BA^{-1}$. This will give

(i)

$$\begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} A & C \\ pB & pD \end{bmatrix} \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} = \left[\begin{array}{c|c} A + pCBA^{-1} & C \\ \hline p^2(D - BA^{-1}C)BA^{-1} & p(D - BA^{-1}C) \end{array}\right]$$

(ii)

$$\begin{bmatrix} I & 0 \\ -Y & I \end{bmatrix} \begin{bmatrix} A & pC \\ B & pD \end{bmatrix} \begin{bmatrix} I & 0 \\ Y & I \end{bmatrix} = \left[\begin{array}{c|c} A + pA^{-1}CB & p^2A^{-1}C(D - BA^{-1}C) \\ \hline B & p(D - BA^{-1}C) \end{array}\right]. \tag{3.5}$$

Iterating these will again give (3.4). We shall also need the next three results in one of our algorithms.

LEMMA 3.6. *Let $C_{r \times n}$ be over $Z/p^m$. Then the following are equivalent:*

(i) *$C$ has $r$ linearly independent rows,*

(ii) *$\rho(C) = r$,*

(iii) *$CC^- = I_r$ for some $C^-$.*

*Proof.* (i)⇔(ii): $C$ has $r$ independent rows ⇔ $C_0$ has $r$ independent rows ⇔ $C_0$ has an $r \times r$ unit minor ⇔ $C$ has an $r \times r$ unit minor ⇔ $\rho(C) = r$.

(ii)⇒(iii): If $\rho(C)=r$, then there is a permutation matrix $K$ such that $CK=[C',C'']$, with $C'$ invertible. Hence $CL=[I_r,0]$ for some unit $L$, and hence $CC^-=I_r$ for some $C^-$. (iii)⇒(i): Clear. ■

COROLLARY 3.7. *If $A\in M_n(Z/p^m)$, then the following are equivalent:*

(i) *the columns of $A$ are linearly independent,*
(ii) $\rho(A)=n$,
(iii) $A^{-1}$ *exists,*
(iv) *the rows of $A$ are linearly independent.*

COROLLARY 3.8. *If $A\in M_n(Z/p^m)$ and* RS$(A)$ *has a basis, then there is a unit matrix $R$ such that*

$$RA=\begin{bmatrix} C \\ 0 \end{bmatrix},$$

*where $C$ is a basis matrix for* RS$(A)$.

*Proof.* Suppose that the rows of $C_{r\times n}$ form a basis for RS$(A)$. Then $A=SC$ and $C=QA$ for some matrices $S$ and $Q$. Hence $(I-QS)C=0$, which by the independence of the rows of $C$ implies that $QS=I_r$. Now by the Lemma 3.6, there is a permutation matrix $K$ such that $QK=[Q_1,Q_2]$, with $Q_1$ invertible. Hence

$$C=QA=QKK^TA=[Q_1,Q_2]\begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

where the rows of $A_2$ are $n-r$ of the rows of $A$. Then

$$\begin{bmatrix} Q_1 & Q_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}=\begin{bmatrix} C \\ A_2 \end{bmatrix}\underset{\text{row}}{\sim}\begin{bmatrix} C \\ 0 \end{bmatrix}$$

as desired. ■

*Method II.* $A_0^{\#}$ *Exists*

Consider $A=A_0+pB$, and suppose

$$Q^{-1}A_0Q=\begin{bmatrix} U_0 & 0 \\ 0 & 0 \end{bmatrix}$$

with $U_0$ invertible. Then

$$Q^{-1}AQ=\left[\begin{array}{c|c} U_0+pL_1 & pL_3 \\ \hline pL_2 & pL_4 \end{array}\right] \qquad \text{for some} \quad L_i.$$

Using (3.5), we may now compute a Fitting decomposition by iteration and hence compute $A^d$.

*Method III. Robert's Method*

If $A^{\#}$ exists, then so does $A_0^{\#}$, and hence Method II applies. An alternative method is the following. Since $A$ is regular, with $\rho(A)=r$, we know that $\mathrm{RS}(A)$ has a basis with $r$ independent rows. Hence by Corollary 3.8, there exists a unit matrix $R$ such that

$$RA=\begin{bmatrix} C \\ 0 \end{bmatrix},$$

in which the rows of $C$ from a basis for $\mathrm{RS}(A)$. Hence

$$RAR^{-1}=\begin{bmatrix} C_1 & C_2 \\ 0 & 0 \end{bmatrix}=A'.$$

Now because $(A')^{\#}$ exists, it follows that [10] $C_1^{\#}$ exists and $C_1C_1^{\#}C_2=C_2$. Consequently,

$$A'\approx\begin{bmatrix} C_1 & 0 \\ 0 & 0 \end{bmatrix}.$$

But $\rho(C_1)=r$ and $C_1$ is $r\times r$, which implies that $C_1^{-1}$ exists and

$$A^{\#}\approx\begin{bmatrix} C_1^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

as desired.

Let us now give three methods dealing with the general case.

*Method IV. Power Methods*

Let $A=A_0+pB$, with $A_0$ over $Z/p$. Since $\mathrm{index}(A)\leqslant mk_0$, we may compute $A^q$ with $q\geqslant mk_0$. For example, $q=mn$ will do. Then $(A^q)^{\#}$ exists and may be computed by Robert's method. $A^d$ is then recovered from $A^{q-1}(A^q)^{\#}$. Alternatively we may compute $A^{k_0} = A_0^{k_0} + p(?)$, in which $(A_0^{k_0})^{\#}$ exists. This time $(A^{k_0})^d$ may be computed using Method II and $A^d$ will then be given by $A^{k_0-1}(A^{k_0})^d$.

In both power methods, as well as in Method I, some powering was used to eliminate part of the $p$-dependence. Let us now turn to an algorithm, which *only* uses an iteration of elementary operations.

*Method V. Roth's Removal Rule*

Consider $A = A_0 + pB$, and suppose that

$$Q^{-1}A_0Q = \begin{bmatrix} U_0 & 0 \\ 0 & \eta_0 \end{bmatrix},$$

with $U_0$ invertible and $\eta_0$ nilpotent. Then

$$Q^{-1}AQ = \begin{bmatrix} U_0 + pL_1 & pL_3 \\ pL_2 & \eta_0 + pL_4 \end{bmatrix}$$

for some $L_i$. Now since $U_0 + pL_1$ is invertible and $\eta_0 + pL_4$ is nilpotent, Corollary 3.5 applies, and

$$A \approx \begin{bmatrix} U' & 0 \\ 0 & N' \end{bmatrix},$$

where $U'$ is invertible and $N'$ is nilpotent. It is now clear that

$$A^d \approx \begin{bmatrix} (U')^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

It should be emphasized that the above methods are algorithmic in nature, and hence will not furnish an explicit expression for $A^d$ in terms of $A_0^d$, $A_1$, etc. The closest we come to an exact formulation is the following.

*Method VI. Block Matrix Method*

We have seen in (1.3) that we may consider a matrix $A \in M_n(Z/p^m)$ as a polynomial in $p$ with matrix entries. We can go one step further and identify

$A(p)$ with $mn\times mn$ block Toeplitz matrix

$$T_A=\begin{bmatrix} A_0 & & & & & \\ A_1 & A_0 & & & \mathbf{0} & \\ A_2 & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ \cdot & & & & \cdot & \\ A_{m-1} & \cdot & \cdot & \cdot & \cdot\; A_1 & A_0 \end{bmatrix}\in M_{mn}(Z/p). \tag{3.6}$$

The map $\phi: A\to T_A$ is easily seen to be a ring as well as a vector space isomorphism, with scalars from $Z/p$. In particular $\phi(p)=J\otimes I$, where

$$J=\begin{bmatrix} 0 & & & & & 0 \\ 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & 0 \end{bmatrix}\quad\text{and}\quad J^m=0.$$

Hence all of the properties of $T_A$ are inherited by $A$. In particular [10, p. 18],

(i) $A^d$ exists$\Leftrightarrow A_0^d$ exists (which is the case),

(ii) $A$ is nilpotent$\Leftrightarrow A_0$ is nilpotent,

(iii) $A^{\#}$ exists$\Leftrightarrow A_0^{\#}$ exists and certain consistency conditions hold [10, p. 19],

(iv) $A^d$ is again a block triangular Toeplitz matrix, and may be found recursively [10, p. 18],

(v) $A^d$ is a polynomial in $A$.

For example, if $m=3$, then $(A_0+A_1p+A_2p^2)^d=B_0+B_1p+B_2p^2$, where for $k\geqslant 3k_0$,

(i) $B_0=A_0^d$,

(ii) $B_1=-B_0A_1B_0+EY_1^{(k)}B_0^{k+1}+B_0^{k+1}Y_1^{(k)}E$ with

$$E=I-A_0A_0^d,\qquad Y_1^{(k)}=\sum_{r=0}^{k-1}A_0^{k-r-1}A_1A_0^r,$$

and

(iii) $B_2 = -(B_1A_1 + B_0A_2)B_0 + EY_2^{(k)}B_0^{k+1} + B_0^{k+1}Y_2^{(k)}E - (A_1B_0 + A_0B_1)Y_1^{(k)}B_0^{k+1} + Z_1^{(k+1)}Y_1^{(k)}E$ with

$$Y_2^{(k)} = \sum_{r=0}^{k-1} A_0^{k-r-1}\left(A_2A_0^r + A_1Y_1^{(r)}\right), \qquad Z_1^{(k+1)} = \sum_{r=0}^{k} B_0^{k-r}B_1B_0^r.$$

Even though $A^d$ has the same Toeplitz structure as $A$, we may in our computation of $A^d$, use elementary row and column operations which do *not* preserve the Toeplitz structure. These operations have no analogue in $M_n(Z/p^m)$ and in essence go outside this ring.

The fact that for a block triangular Toeplitz matrix over $Z/p$, $A^d = A^k$ for some $k$, seems not well known. Of independent interest is the following related fact.

LEMMA 3.9. *If*

$$M = \begin{bmatrix} E_1 & & & \\ & E_2 & & 0 \\ & & \ddots & \\ & ? & & E_k \end{bmatrix}$$

*with $E_i^2 = E_i$, then $M^2 = M$ if and only if $M^{\#}$ and $(I-M)^{\#}$ exist.*

*Proof.* Observe that $M^2 - M = N$ is nilpotent. Hence if $M^{\#}$ and $(I-M)^{\#}$ exist, then $(M^2 - M)^{\#}$ exists, forcing $M^2 - M = 0$. The converse is clear. ∎

Let us now use the above methods on a specific example.

EXAMPLE. Let

$$A = \begin{bmatrix} 2 & 0 & 4 \\ 7 & 5 & 3 \\ 7 & 4 & 0 \end{bmatrix} \in M_3(Z/2^3).$$

Then

$$A^2 = \begin{bmatrix} 0 & 0 & 0 \\ 6 & 5 & 3 \\ 2 & 4 & 0 \end{bmatrix} \quad \text{and} \quad A^3 = A^4 = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 5 & 7 \\ 0 & 4 & 4 \end{bmatrix}.$$

Hence $A^d = A^4 = A^3$. Let us now check this using methods I, II, and V.

*Method I.* Using principal minors we see that $\sigma_1=-1$, $\sigma_2=2$, $\sigma_3=4$. Hence $\Delta_A(\lambda)=\lambda^3+\lambda^2+2\lambda+4=(\lambda^3+\lambda^2)+2(\lambda+2)$ and $\Delta_{A_0}(\lambda)=\lambda^3+\lambda^2$. The desired annihilating polynomial now becomes $(\lambda^3+\lambda^2)^3=\lambda^6(\lambda^3+3\lambda^2+3\lambda+1)=\lambda^6[1-\lambda(-\lambda^2-3\lambda-3)]$ and so $q(\lambda)=-(\lambda^2+3\lambda+3)$. Consequently $A^d=A^6q(A)^7$. Next, $q^2(\lambda)=(\lambda^2+3\lambda+3)^2=\lambda^4-2\lambda^3-\lambda^2+2\lambda+1\equiv 4\lambda+5 \bmod \Delta_A(\lambda)$, and hence $q^4(\lambda)\equiv 1 \bmod \Delta_A$. From this we get $q^7(\lambda)\equiv q^3(\lambda)\equiv -(4\lambda+5)(\lambda^2+3\lambda+3)\equiv 3\lambda^2-3\lambda+1$. Now $\lambda^6\equiv-\lambda^2-2\lambda+4$ and hence $\lambda^6q^7(\lambda)\equiv(-\lambda^2-2\lambda+4)(3\lambda^2-3\lambda+1)\equiv-\lambda^2-2\lambda+4$. This gives $A^d=4I-2A-A^2$, which yields the same matrix.

*Power Method.*

$$A=\begin{bmatrix}0&0&0\\1&1&1\\1&0&0\end{bmatrix}+2\begin{bmatrix}1&0&0\\1&0&1\\1&0&0\end{bmatrix}+4\begin{bmatrix}0&0&1\\1&1&0\\1&1&0\end{bmatrix}.$$

Now

$$A_0^2=\begin{bmatrix}0&0&0\\1&1&1\\1&0&0\end{bmatrix}^2=\begin{bmatrix}0&0&0\\0&1&1\\0&0&0\end{bmatrix}$$

and $\rho(A_0^2)\neq\rho(A_0)$. Thus $A_0$ has no group inverse. We therefore compute $\Delta_{A_0}(\lambda)=\lambda^2(\lambda-1)$ using principal minors, and because $A_0(A_0-I)\neq 0$, we see that $\psi_{A_0}=\Delta_{A_0}$ and $k_0=\operatorname{index}(A_0)=2$. Next compute

$$A^2=\begin{bmatrix}0&0&0\\6&5&3\\2&4&0\end{bmatrix}$$

$$=\begin{bmatrix}0&0&0\\0&1&1\\0&0&0\end{bmatrix}+2\begin{bmatrix}0&0&0\\1&0&1\\0&0&0\end{bmatrix}+4\begin{bmatrix}0&0&0\\1&1&0\\0&1&0\end{bmatrix}.$$

Clearly $A_0^2$ is idempotent. Let us now compute $(A^2)^d$ by elementary operations. First we must reduce $A_0^2$ to canonical form using counters. We shall use two counters, to keep track of the elementary row operations used and their inverses:

$$\left[A_0^2\,\middle|\,I\,\middle|\,I\right]=\Bigg(\left[\begin{array}{ccc|ccc|ccc}0&0&0&1&0&0&1&0&0\\0&1&1&0&1&0&0&1&0\\0&0&0&0&0&1&0&0&1\end{array}\right]$$

$$\sim\begin{array}{ccc|ccc|ccc} & & & \multicolumn{2}{c}{\frown} & & \multicolumn{2}{c}{\frown} & \\ \hline 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}$$

$$\sim\begin{array}{ccc|ccc|ccc} \multicolumn{2}{c}{\overset{-1}{\frown}} & & & & & & \multicolumn{2}{c}{\overset{-1}{\frown}} \\ \hline 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}$$

$$\sim{+1}\Big(\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\sim\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$=\left[\, RA_0^2R^{-1} \mid R \mid R^{-1} \right].$$

Next, we form

$$RA^2R^{-1}=\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ -2 & 5 & 3 \\ 2 & 4 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}=\left[\begin{array}{c|cc} 1 & 0 & 2 \\ \hline 0 & 0 & 0 \\ 4 & 2 & 4 \end{array}\right].$$

We must now iterate to remove the powers of 2. In this case only a couple of steps are needed:

$$\left[\, RA^2R^{-1} \mid I \mid I \right]={+4}\Big(\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 4 & 2 & 4 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\sim\begin{array}{ccc|ccc|ccc} \multicolumn{2}{c}{\overset{+4}{\frown}} & & & & & \multicolumn{2}{c}{\overset{+4}{\frown}} & \\ \hline 1 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 4 & 0 & 1 & 0 & 0 & 1 \end{array}$$

$$\sim\begin{array}{ccc|ccc|ccc} \multicolumn{2}{c}{\overset{-2}{\frown}} & & & & & \multicolumn{2}{c}{\overset{-2}{\frown}} & \\ \hline 1 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 4 & 0 & 1 & 4 & 0 & 1 \end{array}$$

$$\sim_{+2}\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 4 & 0 & 1 & 4 & 0 & 1 \end{array}\right]$$

$$\sim\left[\begin{array}{ccc|ccc|ccc} \overset{+4}{1} & 4 & 0 & 1 & 0 & 2 & \overset{+4}{1} & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 4 & 0 & 1 & 4 & 0 & 1 \end{array}\right]$$

$$\sim_{+4}\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 2 & 1 & 4 & -2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 4 & 0 & 1 & 4 & 0 & 1 \end{array}\right]$$

$$\sim\left[\begin{array}{c|cc|ccc|ccc} 1 & 0 & 0 & 1 & 4 & 2 & 1 & 4 & -2 \\ \cline{1-3} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 4 & 0 & 1 & 4 & 0 & 1 \end{array}\right]$$

$$= [\, SRA^2R^{-1}S^{-1} \mid S \mid S^{-1} \,].$$

Hence

$$(A^2)^d = R^{-1}S^{-1}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} SR = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 5 & 7 \\ 0 & 4 & 4 \end{bmatrix}$$

and $A^d = A(A^2)^d$, which in this case reduces to $(A^d)^2 = (A^2)^d$.

*Removal Rule.* First reduce $A_0$ to its Fitting decomposition using elementary operations and counters:

$$[\, A_0 \mid I \mid I \,] \sim [\, RA_0 \mid R \mid R^{-1} \,]$$

$$\sim\left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\sim_{-1}\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & \cdot & \cdot & 1 & \cdot & \cdot & 1 \\ 1 & 1 & 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot \\ 0 & 0 & 0 & 1 & \cdot & \cdot & \underset{+1}{1} & \cdot & \cdot \end{array}\right]$$

$$\sim\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right]$$

implies

$$
\begin{aligned}
&\left[\,R_1A_0R_1^{-1}\;\middle|\;R_1\;\middle|\;R_1^{-1}\right]\\
&=\circlearrowleft\left[\begin{array}{ccc|ccc|ccc}
0 & 0 & 1 & 0 & 0 & 1 & \overset{\curvearrowright}{0} & 0 & 1\\
2 & 1 & 0 & 0 & 1 & -1 & 1 & 1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0
\end{array}\right]\\
&\sim\left[\begin{array}{ccc|ccc|ccc}
\overset{\curvearrowright}{2} & 1 & 0 & 0 & 1 & -1 & 0 & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}\right]\\
&\sim\left[\begin{array}{ccc|ccc|ccc}
\overset{\overset{-2}{\curvearrowright}}{1} & 2 & 0 & 0 & 1 & -1 & \overset{\overset{-2}{\curvearrowright}}{0} & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}\right]\\
&\sim_{+2}\circlearrowleft\left[\begin{array}{ccc|ccc|ccc}
1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 1 & 1 & -1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}\right]\\
&\sim\left[\begin{array}{ccc|ccc|ccc}
\overset{\overset{-2}{\curvearrowright}}{1} & 0 & 2 & 0 & 1 & 1 & \overset{\overset{-2}{\curvearrowright}}{0} & 0 & 1\\
0 & 0 & 1 & 0 & 0 & 1 & 1 & -1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}\right]\\
&\sim_{+2}\circlearrowleft\left[\begin{array}{c|cc|ccc|ccc}
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1\\
\cline{1-3}
0 & 0 & 1 & 0 & 0 & 1 & 1 & -1 & -2\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}\right]\\
&\sim\left[\begin{array}{c|cc|ccc|ccc}
1 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 1\\
\cline{1-3}
0 & 0 & 1 & 0 & 0 & 1 & 1 & -1 & -2\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{array}\right]\\
&=\left[\,RA_0R^{-1}\;\middle|\;R\;\middle|\;R^{-1}\right].
\end{aligned}
$$

Next we compute

$$
RAR^{-1}=\begin{bmatrix}2&1&1\\0&0&1\\1&0&0\end{bmatrix}\begin{bmatrix}2&0&4\\-1&-3&3\\-1&4&0\end{bmatrix}\begin{bmatrix}0&0&1\\1&-1&-2\\0&1&0\end{bmatrix}=\left[\begin{array}{c|cc}1&2&0\\\hline 4&4&-1\\0&4&2\end{array}\right].
$$

Hence

$$A=1,\quad B=\begin{bmatrix}4\\0\end{bmatrix},\quad C=[2,0],\ \text{and}\ D=\begin{bmatrix}4 & -1\\4 & 2\end{bmatrix}.$$

Now

$$X=[2,0]\left(I+\begin{bmatrix}4 & -1\\4 & 2\end{bmatrix}+\begin{bmatrix}4 & 2\\0 & 0\end{bmatrix}\right)=[2,2],$$

and $BX=0$. Thus

$$\left[\begin{array}{c|cc}1 & 2 & 2\\\hline 0 & 1 & 0\\0 & 0 & 1\end{array}\right]\left[\begin{array}{c|cc}1 & 2 & 0\\\hline 4 & 4 & -1\\0 & 4 & 2\end{array}\right]\left[\begin{array}{c|cc}1 & -2 & -2\\\hline 0 & 1 & 0\\0 & 0 & 1\end{array}\right]=\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 4 & 4 & -1\\0 & 4 & 2\end{array}\right].$$

This time

$$B=\begin{bmatrix}4\\0\end{bmatrix}$$

and

$$Y=\left(I+\begin{bmatrix}4 & -1\\4 & 2\end{bmatrix}+\begin{bmatrix}4 & 2\\0 & 0\end{bmatrix}\right)\begin{bmatrix}4\\0\end{bmatrix}=\begin{bmatrix}4\\0\end{bmatrix}.$$

This gives

$$\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 4 & 1 & 0\\0 & 0 & 1\end{array}\right]\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 4 & 4 & -1\\0 & 4 & 2\end{array}\right]\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 4 & 1 & 0\\0 & 0 & 1\end{array}\right]=\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 0 & 4 & -1\\0 & 4 & 2\end{array}\right].$$

Combining all of these we obtain

$$A^d=\begin{bmatrix}0 & 0 & 1\\1 & -1 & -2\\0 & 1 & 0\end{bmatrix}\left[\begin{array}{c|cc}1 & -2 & -2\\\hline 0 & 1 & 0\\0 & 0 & 1\end{array}\right]\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 4 & 1 & 0\\0 & 0 & 1\end{array}\right]\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 0 & 0 & 0\\0 & 0 & 0\end{array}\right]$$

$$\times\left[\begin{array}{c|cc}1 & 0 & 0\\\hline 4 & 1 & 0\\0 & 0 & 1\end{array}\right]\left[\begin{array}{c|cc}1 & 2 & 2\\\hline 0 & 1 & 0\\0 & 0 & 1\end{array}\right]\begin{bmatrix}2 & 1 & 1\\0 & 0 & 1\\1 & 0 & 0\end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 1 \\ 1 & -1 & -2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 4 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 4 & 5 & -1 \\ 0 & 4 & 4 \end{bmatrix},$$

as desired.

## 4. CANONICAL FORMS

Let us now turn to the question of canonical forms for matrices in $M_n(Z/p^m)$. First of all let us restrict ourselves to elementary row operations. If we embed $M_n(Z/p^m)$ in the ring of $\lambda$-matrices over the field $Z/p$, then we may use the following [6, p. 135].

LEMMA 4.1. *Let $\mathbb{F}$ be a field, and let $A(\lambda) \in M_n(\mathbb{F}[\lambda])$. Then $A(\lambda)$ has a unique $\lambda$-row echelon form defined by*

$$\begin{array}{c} \begin{array}{ccccccccccc} & i_1 & & & i_2 & & & i_3 & & i_r & \end{array} \\ \left[ \begin{array}{ccccccccccc} 0 & a_1(\lambda) & & & & & & & & & \\ 0 & 0 & \cdots & 0 & a_2(\lambda) & & & & & & \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & a_3(\lambda) & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & a_r(\lambda) & \cdots \\ \hline \multicolumn{11}{c}{0} \end{array} \right] \end{array}$$

*such that*

(i) *the zero rows are at the bottom,*

(ii) *the first nonzero entry in each row (called a pivot) is monic and has zeros below it,*

(iii) *if $i<j$, then the row position of $a_i$ is above that of $a_j$,*

(iv) *the elements above $a_i$ are either absent, zero, or of degree less than $\partial a_i$.*

The uniqueness of this form, except for the case where $\det A(\lambda) \neq 0$, does not seem to be well known. If we interchange the rows so that the pivots $a_i(\lambda)$ fall on the diagonal, then the unique canonical form is called the Hermite normal form of $A(\lambda)$.

If we apply the $\lambda$-row echelon form to a matrix $A(p)$ over $Z/p^m$, then the outcome is no longer unique due to the ambiguity of the factor $p^m$. For example if $p^3=0$, then

$$\begin{bmatrix} p & p & 1 \\ 0 & p^2 & p \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} p & p & 1+p^2 \\ 0 & p^2 & p \\ 0 & 0 & 0 \end{bmatrix},$$

both of which are in echelon form. In order to get around this problem, one might try to row-reduce $A(p)$ using the *order* of the entries in $A(p)$ in the obvious manner. In the first nonzero column, select an element of minimal order. Push this element to the first row and reduce it to the form $p^k$, by multiplying through by a unit. Then sweep out the rest of column one. Next, delete row one and repeat. After this process has terminated (either because there are no more nonzero rows, or because we have reached the last row), we divide the pivots into the entries above them, ensuring that they have a degree (as well as an order) which is smaller than that of the pivot below them. Again this form will not be unique. For example, if $p^3=0$, then

$$\begin{bmatrix} 1 & p \\ p & 1 \end{bmatrix} \begin{bmatrix} p^2 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} p^2 & 1 \\ 0 & p \end{bmatrix},$$

with again the trouble coming from the term $p^3$. To get around this difficulty, we may use the Fuller canonical form [5] for $A \in M_n[Z/p^m)$, which is characterized by the following relationship between a diagonal entry and its row and column containing it:

$$A \underset{\text{row}}{\sim} F_A = \left[\begin{array}{c|c|c} & c & \\ \hline a & d & b \\ \hline & e & \end{array}\right],$$

with

(i) $d=p^k$ for some $k$,
(ii) $\delta a \geqslant \delta d$, $\delta b > \delta d$,
(iii) $c=0$ or $\partial c < \partial d$,
(iv) $e=0$ or $\partial e < \partial d$.

This canonical is unique and may be obtained by selecting a rightmost element with minimal order, pushing it to the diagonal, reducing it to $p^k$, and sweeping out the rest of its column. After deleting the pivot row and column the process is repeated. When the process terminates, the diagonal pivots may then be used to reduce the *degrees* of all nonzero elements in the columns of each particular pivot.

For $\lambda$-matrices over a field $\mathbb{F}$, we may obtain a similar canonical form, using *degrees* instead of orders. The main reason for using degrees is that there is *no* g.c.d. algorithm which uses orders. Indeed, if $A\in M_n(\mathbb{F}[\lambda])$, then

$$A(\lambda)\sim\left[\begin{array}{c|c|c} & c & \\ \hline a & d & b \\ \hline & e & \end{array}\right],$$

where

(i) $d$ is monic
(ii) $\partial d \geqslant \partial a,\ \partial d > \partial b,$
(iii) $c=0$ or $\partial c < \partial d,$
(iv) $e=0$ or $\partial e < \partial d.$

This form is unique within ordering of the diagonal entries. That is, two canonical forms are identical if corresponding diagonal elements have the same degree. For example, in

$$\begin{bmatrix} 1 & \lambda^2-1 \\ \lambda^2+1 & \lambda^4 \end{bmatrix}\begin{bmatrix} \lambda^4 & 0 \\ -1-\lambda^2 & 1 \end{bmatrix}=\begin{bmatrix} 1 & \lambda^2-1 \\ 0 & \lambda^4 \end{bmatrix}$$

the last two matrices are in canonical form, but they are not equal.

The reason for considering the Fuller form is that it can be used to give another algorithm for computing $A^d$. Indeed, we shall now show that if $A$ is regular, then its Fuller form is idempotent and can be used to construct bases for $R(A)$ and $N(A)$, exactly as in the field case.

THEOREM 4.2. *Let $A\in M_n(Z/p^m)$. The following are equivalent:*

(i) $\mathrm{RS}(A)$ *has a basis,*
(ii) $F_A^2=F_A,$
(iii) $A$ *is unit regular,*
(iv) $A$ *is regular,*
(v) $R(A)$ *has a basis.*

*Proof.* (i)$\Leftrightarrow$(ii): By Corollary 3.7, there is a unit $R$ such that

$$RA=\begin{bmatrix} C \\ 0 \end{bmatrix},$$

where the rows of $C$ form a basis for RS$(A)$. In particular, the first row of $C$ is independent and must therefore contain a unit. Selecting the *rightmost* unit in row one we see that

$$A \underset{\text{row}}{\sim} \left[\begin{array}{ccc} & 1 & \\ ? & 0 & ? \\ & 0 & \\ & \vdots & \\ \hline & 0 & \end{array}\right]_{r},$$

in which the new nonzero rows again form a basis. Repeating this argument with row 2, we see that

$$A \underset{\text{row}}{\sim} \begin{array}{c} \begin{array}{cccc} i_2 & i_r & i_1 & i_3 \end{array} \\ \left[\begin{array}{cccc} 0 & 0 & 1 & 0 \\ 1 & \vdots & 0 & \vdots \\ 0 & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & 1 \\ \vdots & \vdots & \vdots & 0 \\ \vdots & 0 & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ \hline & & 0 & \end{array}\right] \end{array} \underset{r\,\text{row}}{\sim} \begin{array}{c} \begin{array}{ccccccccc} & i_3 & & i_2 & & i_1 & & & \end{array} \\ \left[\begin{array}{ccccccccc} 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & \cdots & 0 \\ \vdots & 1 & & \vdots & & \vdots & & & \vdots \\ \vdots & 0 & \ddots & 0 & & \vdots & & & \vdots \\ \vdots & \vdots & & 1 & & \vdots & & & \vdots \\ \vdots & \vdots & & 0 & \ddots & 0 & & & \vdots \\ \vdots & \vdots & & \vdots & & 1 & & & \vdots \\ \vdots & \vdots & & \vdots & & 0 & 0 & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & \cdots & 0 \end{array}\right] \end{array}$$

$$=F,$$

where the following facts hold:

(1) On the diagonal there are $r$ pivots of 1 and $n-r$ zeros. That is, $f_{ii}=1$ if $i \in I=\{i_1, i_2, \dots, i_r\}$ and $f_{ii}=0$ for $i \notin I$.

(2) Above and below the unit pivots we have zeros. That is, $f_{ij}=0$ if $j \in I$ and $i \neq j$.

(3) If $f_{ii}=0$, then the rest of the row is zero.

(4) $f_{ij}=0$ for $i>r$.

It is now easily seen that $F$ is indeed in Fuller form and that $F^2=F$. In fact, for $i \notin I$,

$$(F^2)_{ij}=\sum_k f_{ik} f_{kj}=0,$$

while if $i\in I$, then

$$(F^2)_{ij}=\sum_{k\in I} f_{ik}f_{kj}+\sum_{k\notin I} f_{ik}f_{kj}=\sum_{k\in I} f_{ik}f_{kj}=f_{ij}.$$

(ii)⇔(iii): $RA=F^2=F\Rightarrow ARA=A$, and $A$ is unit regular.
(iii)⇒(iv): Clear.
(iv)⇒(i): Done in (2.2) (iii).
(i)⇒(v): Use transposes. ■

Corollary 4.3. *Let $A$ be regular of rank $r$. Then*

(i) *the columns $a_{i_j}$ in $A$ corresponding to the unit pivots in the Fuller normal form yield a basis for $R(A)$,*

(ii) *the nonzero columns in $I-F$ yield a basis for $N(A)$,*

(iii) *any set of $r$ linearly independent rows (columns) forms a basis for* RS$(A)$ $(R(A))$.

*Proof.* (i): Let

$$RA=F_A=\begin{bmatrix} C\\ 0\end{bmatrix},$$

with $R$ invertible, and with $\mathbf{e}_k$ in column $i_k$ of $F$, $k=1,2,\dots,r$. Let $B=[\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r}]$. Then

$$RB=\begin{bmatrix} I_r\\ 0\end{bmatrix}\quad\text{and}\quad R^{-1}=[B,?].$$

Hence

$$A=R^{-1}F=R^{-1}\begin{bmatrix} C\\ 0\end{bmatrix}=[B,?]\begin{bmatrix} C\\ 0\end{bmatrix}=BC.$$

This means that $A$ again admits a full-rank factorization and that the columns of $B$ yield a basis for $R(A)$.

(ii): Since $ARA=A$, we know that $N(A)=R(I-RA)=R(I-F)$. That is, the $n-r$ nonzero columns of $I-F$ span $N(A)$. They are also independent, because if $(I-F)\mathbf{x}=\mathbf{0}$ with $x_i=0$ for $i\in I=\{i_1,\dots,i_r\}$, then $\mathbf{x}=F\mathbf{x}$, which implies that $x_i=0$ for $i\notin I$ also. Hence $\mathbf{x}=\mathbf{0}$.

(iii): Let $\rho(A)=r$, and suppose that $C$ contains $r$ independent rows of $A$. Then

$$R_1A=\begin{bmatrix} C \\ ? \end{bmatrix}$$

for some unit matrix $R$. By Lemma 3.5, there is a permutation matrix $K$ such that

$$R_1AK=\begin{bmatrix} C_1 & C_2 \\ ? & X \end{bmatrix}$$

with $C_1$ invertible. Thus

$$RAK=\begin{bmatrix} C_1 & C_2 \\ 0 & X \end{bmatrix}$$

for some unit $R$. Now because $\rho(A)=r$, it follows that $X=pY$, and since $A$ is regular, we must have $Y=0$. Consequently

$$RA=\begin{bmatrix} C \\ 0 \end{bmatrix},$$

which shows that the rows of $C$ also span RS$(A)$.

Suppose now that $A^{\#}$ exists. Then $\mathbb{R}^n=R(A)\dot{+}N(A)$ and hence we may compute a basis matrix $Q=[Q_1,Q_2]$ for $\mathcal{R}^n$, where $Q_1$ is a basis for $R(A)$ and $Q_2$ is a basis matrix for $N(A)$. Moreover

$$AQ=Q\begin{bmatrix} U & 0 \\ 0 & 0 \end{bmatrix}$$

for some unit matrix $U$, and $A^{\#}$ may be computed. In general, this may be used to replace Robert's algorithm in the power method. ■

Let us conclude with a brief examination of row or column equivalence over $Z/p^m$. It is well known [4] that there exist invertible $R$ and $K$ such that

$$RAK=\begin{bmatrix} I_{r_0} & & & & 0 \\ & pI_{r_1} & & & \\ & & p^2I_{r_2} & & \\ & & & \ddots & \\ 0 & & & & p^mI_{r_m} \end{bmatrix}=S_A \tag{4.1}$$

for some $r_i \geqslant 0$. It should be pointed out here that this normal form is really an immediate consequence of the Smith normal form for $A(\lambda)$. Indeed, if

$$S_A(\lambda)=\begin{bmatrix} s_1(\lambda) & & & \\ & s_2(\lambda) & & 0 \\ & & \ddots & \\ & 0 & & s_n(\lambda) \end{bmatrix}$$

is the Smith normal form of $A(\lambda)$, then $s_i(\lambda)=\alpha_0^{(i)}+\alpha_1^{(i)}\lambda+\cdots+\alpha_{m-1}^{(i)}\lambda^{m-1}$. Suppose that $\alpha_0^{(i)}=0=\cdots=\alpha_{k-1}^{(i)}\neq\alpha_k^{(i)}$. Then $s_i(p)=p^k u_i$ for some unit $u_i$. This means that for every $i=1,2,\ldots,n$, $s_i$ can be row reduced to $p^{k_i}$ for some $k_i$. A final permutation yields the desired normal form.

From $S_A$ it is clear that indeed,

(1) $A$ is regular $\Leftrightarrow A\sim\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$
$\Leftrightarrow R(A)$ has basis
$\Leftrightarrow$ RS$(A)$ has a basis
$\Leftrightarrow A$ is unit regular,

(2) most of the results derived by Roth [15] dealing with solutions to the matrix equations $A(\lambda)X(\lambda)-Y(\lambda)B(\lambda)=C(\lambda)$ may be used to derive solutions for the corresponding matrix equations over $Z/p^m$ [4].

Let us close with some open questions:

(1) Are there canonical forms for $A\in M_n(Z/p^m)$ under similarity, other than the Fitting decomposition?

(2) If every idempotent matrix $E$ is similar to $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ for some $r\geqslant 0$, does this suffice for every nonzero idempotent to have a unit entry?

(3) Can the Souriau-Frame algorithm be modified to $M_n(Z/p^m)$?

(4) Can $\operatorname{adj}(\lambda I-A(p))$ be used to compute $A^d$?

## REFERENCES

1 A. Ben Israel and T. N. E. Greville, *Generalized Inverses, Theory and Applications*, Wiley, New York, 1974.

2 S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations*, Pitman, New York, 1979.

3 M. P. Drazin, Pseudo-inverses in Associated Rings and Semigroups, *Amer. Math. Monthly* 65:506–514 (1958).

4 T. P. Donovan, Certain Matrix Congruences mod $p^n$, *Ann. Mat. Pura Appl. IV* 65:193–214 (1977).

5 L. E. Fuller, A canonical set of matrices over a principal ideal ring modulo $m$, *Canad. J. Math.* 7:54–58 (1955).

6 F. R. Gantmacher, *The Theory of Matrices*, Vol. 1, Chelsea, New York, 1960.

7 R. E. Hartwig, A note on periodic matrices, *J. Industrial Soc.* 27, part 1:51–55. (1977).

8 R. E. Hartwig, Drazin inverses in cryptography, submitted for publication.

9 R. E. Hartwig, From Schur to Jordan, to appear.

10 R. E. Hartwig and J. Shoaf, Group inverses and Drazin inverses of bidiagonal and triangular Toeplitz matrices, *J. Austral. Math. Soc. Ser. A* 24:10–34 (1977).

11 I. Kaplansky, *Fields and Rings*, 2nd ed., Univ. of Chicago Press, Chicago, 1965.

12 J. Levine and R. E. Hartwig, Applications of the Drazin inverse to the Hill cryptographic system, I, *Cryptologia* 4:71–83 (1980).

13 N. H. McCoy, *Rings and Ideals*, Carus Monograph No. 8, Buffalo, 1948.

14 B. R. McDonald, *Finite Rings with Identity*, M. Dekker, New York, 1974.

15 W. E. Roth, The equations $AX-YB=C$ and $AX-XB=C$ in Matrices, *Proc. Amer. Math. Soc.* 3:392–396 (1952).

*Received 7 April 1980; revised 17 June 1980*