The Drazin Inverse as a Gradient

Richard Gabriel
Center for Mathematical Statistics
Academy of Science
Bucharest, Roumania

and

Robert E. Hartwig
Mathematics Department
North Carolina State University
Raleigh, North Carolina 27695-8205

Submitted by Hans Schneider

ABSTRACT

The coefficients in the expansion of $\operatorname{adj}(\lambda I - A)$ are expressed as gradients, and some new representations are given for the Drazin inverse of a matrix over an arbitrary field. These results are then combined to express the Drazin inverse as a gradient of a function of the entries of the matrix.

1. INTRODUCTION

If $X = [x_{ij}]$ is an $m \times n$ matrix over a field $\mathbb{F}$, and $f(X) = f(x_{11}, \dots, x_{mn})$ is a function from $\mathbb{F}_{m\times n}$ into $\mathbb{F}$, depending on the entries of X, then the gradient $\nabla_X f(X)$ is defined to be the $m \times n$ matrix $(\nabla_X f)_{ij} = (\partial/\partial x_{ij})f$. For example, if x is a column, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, then $\nabla_x f(\mathbf{x}) = [\partial f/\partial x_1, \dots, \partial f/\partial x_n]^T$. Suppose that $A \in \mathbb{F}_{n\times n}$ has determinant $|A|$ and cofactors A_{ij}. Then, considering the entries a_{ij} as independent variables, we may write $(\partial/\partial a_{ij})|A| = A_{ij}$. That is,

$$\operatorname{adj}(A)^T = \nabla_A |A|. \tag{1.1}$$

52 Vanderbilt Ave., New York, NY 10017 0024-3795/84/$3.00

For an invertible matrix A, this may be used to rewrite Cayley's formula $A^{-1} = \mathrm{adj}(A)/|A|$ in gradient form:

$$(A^{-1})^T = \frac{1}{|A|}\mathrm{adj}(A)^T = [\nabla_X \ln|X|]_{X=A}. \tag{1.2}$$

In this, the potential function $\ln|X|$ is formally defined by

$$\frac{\partial}{\partial x_{ij}}\ln f(X) = \frac{1}{f(X)}\frac{\partial}{\partial x_{ij}}f(X).$$

Our aim will be to generalize this result to singular $n \times n$ matrices, by finding a potential function w for the Drazin inverse. That is, we shall construct a function $w = w(x_{11},\dots,x_{nn})$ in the form of $\ln\det W(X)$, so that

$$(A^d)^T = \nabla_X w(X)]_{X=A} = \nabla_X \ln\det W(X)]_{X=A} \tag{1.3}$$

for some suitable matrix $W(X)$.

Our strategy will be to obtain a suitable representation for X^d in terms of coefficients X_i in the expansion of $\mathrm{adj}(\lambda I - X)$, and then to derive the potential function for these coefficients. As always, we shall denote the set of $m \times n$ matrices over $\mathbb{F}$ by $\mathbb{F}_{m\times n}$, and use rank$(\cdot)$, $(\cdot)^T$, and det$(\cdot)$ or $|\cdot|$ to denote the rank, transpose, and determinant respectively. We shall further use $\mathbf{e}_i$ for the unit vector $[0,\dots,1,0,\dots,0]^T$ and shall write $\mathbb{D}_\lambda$ to denote formal differentiation with respect to λ. Some knowledge of the elementary properties of the Drazin inverse X^d of a matrix X [1] will be assumed, and we shall shorten the term "generalized inverse" to g-inverse.

2. THE ADJOINT EXPANSION

Let $X \in \mathbb{F}_{n\times n}$, and suppose that its characteristic and minimal polynomials are given by

$$\begin{aligned}\Delta(X,\lambda) &= |\lambda I - X| = \lambda^k(x_k + x_{k+1}\lambda + \cdots + \lambda^{n-k}) = \lambda^k\Delta_0(\lambda),\\ \psi(X,\lambda) &= \lambda^l(\tilde{x}_l + \tilde{x}_{l+1}\lambda + \cdots + \lambda^{m-l}),\end{aligned} \tag{2.1}$$

where $x_k \neq 0 \neq \tilde{x}_l$ and $k \geqslant l \geqslant 0$.

The integer k is sometimes called the *zero order* of X, while the integer $\rho = n - k$ is usually called the *core rank* of X. In addition the exponent $l = l(X)$ is often referred to as the *index* of X. Now if

$$X = Q\begin{bmatrix} U & 0 \\ 0 & \eta \end{bmatrix}Q^{-1} \tag{2.2}$$

is Fitting's decomposition of X, with U invertible and η nilpotent, in $\mathbb{F}_{k\times k}$, then the *core* C_X of X is given by

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

while the nilpotent part N_X is equal to

$$N_X = Q\begin{bmatrix} 0 & 0 \\ 0 & \eta \end{bmatrix}Q^{-1}.$$

Clearly $\Delta_0(\lambda) = \Delta(U,\lambda)$, $\rho = n - k = \text{rank}(C_X) = \text{rank}(U)$, and $l(\eta) = l(X)$. We shall show that in terms of these coefficients, the potential matrix $W(X)$ of (1.3) can be expressed as

$$W(X) = \frac{1}{x_k}\begin{bmatrix} x_k^2 & x_k x_{k-1} & \cdots & x_k x_0 \\ x_{k+1} & x_k & & \\ \vdots & \vdots & \ddots & \bigcirc \\ x_{2k} & x_{2k-1} & \cdots & x_k \end{bmatrix}_{(k+1)\times(k+1)}. \tag{2.3}$$

For algebraic purposes, it is often convenient to write the characteristic polynomial of X in the form $\Delta(X,\lambda) = \lambda^n - \sigma_1\lambda^{n-1} + \sigma_2\lambda^{n-2} \cdots (-1)^{n-r}\sigma_{n-r}\lambda^r$. Indeed, the coefficients σ_r can be expressed for all $r = 0,1,\ldots,n$ as

$$\sigma_r = (-1)^r x_{n-r} = \sum_{\alpha} |X_\alpha^\alpha|, \qquad \alpha = (\alpha_1,\ldots,\alpha_r)^T, \tag{2.4}$$

which represents the sum of all $r\times r$ principal minors of X. Here X_β^α denotes the submatrix of X generated by rows $(\alpha_1,\ldots,\alpha_r)$ and columns $(\beta_1,\beta_2,\ldots,\beta_s)$.

Suppose further that the λ-adjoint has been expanded as

$$\operatorname{adj}(\lambda I - X) = X_0 + X_1\lambda + \cdots + I\lambda^{n-1}, \tag{2.5}$$

where the X_r are the so called "adjoint coefficients" [4]. It is easily seen from (2.1) and (2.5) that

$$X_0 = X_1 = \cdots = X_{k-l-1} = 0 \neq X_{k-l}, \tag{2.6}$$

and that

$$\begin{gathered} XX_i = X_{i-1} - x_i I, \qquad i = 1,2,\dots,n-1, \\ X_{-1} = 0 = X_n. \end{gathered} \tag{2.7}$$

Hence $X^i X_{k-1} = X_{k-i-1}$, $i = 0,1,\dots,l-1$, from which we see that $(X_{k-l},\dots,X_{k-1})$ are linearly independent, even if X is nilpotent. It is well known that the recurrence relation (2.7) may be solved to yield

$$X_i = p_i(X), \qquad i = -1,0,1,\dots,n-1, \tag{2.8}$$

where the "adjoint polynomials" are given by

$$p_i(\lambda) = x_{i+1} + x_{i+2}\lambda + \cdots + x_n\lambda^{n-i-1}, \tag{2.9}$$

$i = -1,0,1,\dots,n-1$. In particular $p_{-1}(\lambda) = \Delta(X,\lambda)$, $p_{k-1}(\lambda) = \Delta(U,\lambda)$, $p_k(\lambda) = x_{k+1} + x_{k+2}\lambda + \cdots + \lambda^{n-k-1}$, and $p_{k-i}(\lambda) = \lambda^{i-1}p_{k-1}(\lambda)$, $i = 1,2,\dots,l$. From this it is easily seen that if $Y = Q^{-1}XQ$ then $Y_i = Q^{-1}X_iQ$. Moreover, using the Fitting decomposition, it follows that

$$(X_{k-l}, X_{k-l+1},\dots, X_{k+(m-l+1)}) \tag{2.10}$$

forms a basis for the polynomial space spanned by $(I, X, X^2,\dots)$. It should further be noted that while the matrices $(X_{k-l},\dots,X_{k-2})$ are all nilpotent, the matrix $X_{k-1} = p_{k-1}(X) = \Delta(U,X)$ has a group inverse [4]. By analogy to (2.4) we may rewrite the λ-adjoint as

$$\operatorname{adj}(\lambda I - X) = \Sigma_1\lambda^{n-1} - \Sigma_2\lambda^{n-2} + \cdots + (-1)^{n-1}\Sigma_n, \tag{2.11}$$

where the signed coefficients are given by

$$\Sigma_r = (-1)^{r-1} X_{n-r}, \qquad r = 0, 1, \dots, n. \tag{2.12}$$

The dual notation of (2.4) and (2.12) is necessary if one wants to avoid cumbersome and often unwieldy minus signs. In theoretical considerations it will usually be easier to use x_r and X_r, while when dealing with adjoints and principal minors, it will be more convenient to use σ_r and Σ_r. For example, $\Sigma_n = (-1)^{n-1} X_0 = \mathrm{adj}(X)$.

It will be demonstrated shortly that the entries in Σ_r are in fact representable as sums of algebraic complements of a_{ij} in the $r \times r$ principal minors of A.

From (2.4) we see that the coefficients x_i are *fixed* multinomial functions (that is, sums of products) of the entries $x_{11}, \dots, x_{nn}$. Likewise the coefficients X_i in (2.5) are *fixed* matrix valued functions of the entries $x_{11}, \dots, x_{nn}$. The spectral indices k and l, in turn, are unique functions of the x_{ij}, via

$$k = \min_i \{x_i \neq 0\}$$

$$l = \min_i \{X^i X_{k-1} = 0\}.$$

For a particular matrix A, we may calculate the coefficients a_i and A_i in the characteristic and adjoint expansions (2.1) and (2.5), respectively, from $a_i = x_i(A)$ and $A_i = X_i(A)$.

3. ANALYTIC INVERSES

A fundamental problem in the theory of generalized inverses is to find coefficients $\zeta_i = \zeta_i(X)$ so that the polynomial

$$\zeta(X) = \zeta_0 I + \zeta_1 X + \cdots + \zeta_{n-1} X^{n-1}$$

represents a generalized inverse of X. That is, $\zeta(X) = X^{-1}$, whenever X^{-1} exists. One obvious way of solving this problem is via the Cayley-Hamilton theorem, which says that if X^{-1} exists then $k = 0$, and one may take

$$\zeta(\lambda) = -\frac{1}{x_0}\left[x_1 + x_2\lambda + \cdots + \lambda^{n-1}\right]. \tag{3.1}$$

For singular X, with $\Delta(X,\lambda)$ as in (2.1), it is clear that we may construct the polynomial

$$\begin{aligned}\zeta(\lambda) &= -\frac{1}{x_k}\left[x_{k+1}+x_{k+2}\lambda+\cdots+\lambda^{n-k-1}\right]\\ &= -\frac{1}{x_k}p_k(\lambda)=\frac{1}{\lambda}\left[1-\frac{\Delta(U,\lambda)}{x_k}\right],\end{aligned} \tag{3.2}$$

with the associated generalized inverse

$$\tilde{X} = -\frac{1}{x_k}p_k(X) = -\frac{X_k}{x_k}. \tag{3.3}$$

This matrix has the block form

$$\tilde{x} = Q\begin{bmatrix} U^{-1} & 0 \\ 0 & -p_k(\eta)/x_k \end{bmatrix}Q^{-1},$$

and was called the principal analytic inverse in [2]. It will be invertible precisely when $x_{k+1}\neq 0$.

More generally we may consider the family of analytic spectral inverses [2]

$$\zeta(X) = -\frac{X_k}{x_k}+\sum_{i=1}^{l}\zeta_i X_{k-i} \tag{3.4}$$

for some ζ_i. These g-inverses not only commute with X, but also satisfy the range condition

$$X^{l+1}\zeta = X^l. \tag{3.5}$$

Indeed, since $0 = X^l X_{k-1} = X^l(x_k I + XX_k)$, it follows that $\tilde{X}$ solves (3.5), after which the identities $X^l X_{k-i}=0$, $i=1,2,\dots,l$, ensure that the $\zeta(X)$ have the same property. We shall now show that the Drazin inverse X^d is also a member of this family of analytic inverses (3.4). To do this we shall have to introduce the following notation. If $q(\lambda)=q_0+q_1\lambda+\cdots+\lambda^N$, $q_0\neq 0$, is a polynomial in $\mathbb{F}[\lambda]$, then we may define the coefficients r_i by the determinants

$$r_0 = q_0^{-1}, \qquad r_i = (-1)^i q_0^{-i-1} \begin{vmatrix} q_1 & q_0 & & \bigcirc \\ q_2 & q_1 & \ddots & \\ \vdots & \vdots & \ddots & q_0 \\ q_i & q_{i-1} & \cdots & q_2 \end{vmatrix}_{i\times i}, \qquad i = 1,2,\ldots. \tag{3.6}$$

These coefficients satisfy the matrix equation [4]

$$\begin{bmatrix} q_0 & & & \\ q_1 & q_0 & \bigcirc & \\ \vdots & & \ddots & \\ q_i & \cdots & q_1 & q_0 \end{bmatrix} \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_i \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

for all $i = 1,2,3,\ldots$.

Consequently we may think of the coefficients r_i as being those of the formal power series

$$r(\lambda) = \frac{1}{q(\lambda)} = \sum_{i=0}^{\infty} r_i \lambda^i.$$

Moreover, $q(\lambda)[r_0 + r_1\lambda + \cdots + r_l\lambda^l] = 1 - \lambda^{l+1}s(\lambda)$ for some polynomial $s(\lambda)$.

Consider now the formal differentiation operator $\mathbb{D}_\lambda$. If $\operatorname{char}\mathbb{F} = 0$, then $(1/i!)[\mathbb{D}_\lambda^i(p_0 + p_1\lambda + \cdots p_k\lambda^m)]_{\lambda=0} = p_i$. In other words the operator $(1/i!)\mathbb{D}_\lambda^i(\cdot)]_{\lambda=0}$ extracts the coefficient of λ^i from the polynomial $p(\lambda)$. For $\operatorname{char}\mathbb{F} \neq 0$, we shall use the same notation for this operation.

Moreover we may define

$$\frac{1}{t!}\left[\mathbb{D}_\lambda^t\left(\frac{p(\lambda)}{q(\lambda)}\right)\right]_{\lambda=0} = \sum_{i=0}^{t} p_i r_{t-i}$$

to mean the coefficient of λ^t in the formal product of

$$\frac{1}{q(\lambda)} \cdot p(\lambda) = r(\lambda)p(\lambda).$$

Of course, when char$\mathbb{F}=0$, this is precisely what one would obtain by using the quotient rule on the rational function $p(\lambda)/q(\lambda)$.

Now consider the formal expansion $\Delta(U,\lambda)^{-1}=\Sigma_0^\infty \chi_i\lambda^i$, where the $\chi_i=(1/i!)[\mathbb{D}_\lambda^i(1/\Delta(U,\lambda))]_{\lambda=0}$ are explicitly given by

$$\chi_0=x_k^{-1},\qquad \chi_i=\frac{(-1)^i}{x_k^{i+1}}\begin{vmatrix} x_{k+1} & x_k & & \bigcirc \\ x_{k+2} & x_{k+1} & & \\ \vdots & \vdots & \ddots & x_k \\ x_{k+i} & x_{k+i-1} & \cdots & x_{k+1}\end{vmatrix},\qquad i=1,2,\dots. \tag{3.7}$$

Also let $\Delta^-(U,\lambda)=\chi_0+\chi_1\lambda+\cdots+\chi_l\lambda^l$. Then $\Delta(U,\lambda)\Delta^-(U,\lambda)=1-\lambda^{l+1}s(\lambda)$ for some polynomial $s(\lambda)$, or more formally $\Delta^-(U,\lambda)\equiv\Delta^{-1}(U,\lambda)$ mod λ^{l+1}. Now set $d(\lambda)=[1-\Delta(U,\lambda)\Delta^-(U,\lambda)]/\lambda=\lambda^l s(\lambda)$. Then we have

THEOREM 1. *Let $X\in\mathbb{F}_{n\times n}$ have characteristic polynomial as in* (2.1), *and let $d(\lambda)$ be defined as above. Then*

$$X^d=d(X). \tag{3.8}$$

Proof. Using (2.2), we see that $d(U)=U^{-1}\cdot[I-\Delta(U,U)\Delta^-(U,U)]=U^{-1}$, while $d(\eta)=\eta^l s(\eta)=0$, as desired. ■

We may rewrite $d(\lambda)$ in a form from which its spectral inverse character is more obvious. We have

$$\begin{aligned} d(\lambda)&=\frac{1}{\lambda}\left[1-\frac{\Delta(U,\lambda)}{x_k}\right]+\Delta(U,\lambda)\frac{1-x_k\Delta^-(U,\lambda)}{x_k\lambda}\\ &=-\frac{p_k(\lambda)}{x_k}-\Delta(U,\lambda)\left[\chi_1+\chi_2\lambda+\cdots+\chi_l\lambda^{l-1}\right]\end{aligned}$$

and hence

$$d(\lambda)=-\left[\chi_0p_k(\lambda)+\chi_1p_{k-1}(\lambda)+\cdots+\chi_lp_{k-l}(\lambda)\right]. \tag{3.9}$$

Consequently

$$X^d=-\sum_{i=0}^{l}\chi_iX_{k-i}=\tilde{X}-\sum_{i=1}^{l}\chi_iX_{k-i}. \tag{3.10}$$

It should be remarked here that we may in $\Delta^{-}(U,\lambda)$ and in (3.9) replace l by any t such that $l \leqslant t \leqslant k$. This does not affect (3.10), since the X_{k-i} vanish for $i > l$, and the coefficients χ_i only depend on k.

Let us now proceed and develop the potential function for the adjoint coefficients X_r. This will subsequently be combined with (3.10) to give the potential function for A^d.

4. THE POTENTIAL FUNCTIONS

THEOREM 2

Let x_i and X_i be the coefficients of λ in $\Delta(X,\lambda)$ and $\operatorname{adj}(\lambda I - X)$. *Then*

$$X_r^T = -\nabla_X x_r, \qquad r = 0,1,\ldots,n-1. \tag{4.1}$$

Proof. Consider $\lambda I - X$, where λ is an indeterminate independent of the entries x_{ij}. Then from (1.1) we see that

$$\operatorname{adj}(\lambda I - X)^T = \nabla_{\lambda I - X}\Delta(X,\lambda) = -\nabla_X \Delta(X,\lambda), \tag{4.2}$$

since $\partial(\lambda\delta_{ij} - x_{ij})/\partial x_{ij} = -1$. Hence $\sum_{i=0}^{n-1} X_i \lambda^i = -\sum_{i=0}^{n-1}\nabla_X x_i \lambda^i$. Equating powers of λ^i yields the desired results. ■

Dually this result may be written as

$$\Sigma_r = \nabla_X \sigma_r.$$

When $r = 0$, it is easily seen that (4.1) reduces to (1.1), since $X_0 = (-1)^{n-1}\operatorname{adj}(X)$. Similarly, for $r = 1$ we obtain $I = \nabla_X \operatorname{Tr}(X)$.

Our first consequence of Theorem 1, is the following nontrivial characterization of the entries in the signed adjoint coefficient matrices Σ_r.

COROLLARY 1.

$$(\Sigma_r)_{ji} = \sum_{\alpha} \frac{\partial}{\partial x_{ij}} |X_\alpha^\alpha|, \qquad \alpha = (\alpha_1,\ldots,\alpha_r)^T. \tag{4.3}$$

Proof. First of all, it should be noted that $(\partial/\partial x_{ij})|X_\alpha^\alpha|$ is the algebraic complement of a_{ij} in the principal minors $|X_\alpha^\alpha|$. Consequently its value is zero when i or j do not appear in $\{\alpha_1,\dots,\alpha_r\}$. Now on using (2.4) we see that

$$\sum_{\boldsymbol{\alpha}} \frac{\partial}{\partial x_{ij}}|X^{\boldsymbol{\alpha}}_{\boldsymbol{\alpha}}| = \frac{\partial}{\partial x_{ij}} \sum_{\boldsymbol{\alpha}} |X^{\boldsymbol{\alpha}}_{\boldsymbol{\alpha}}| = \frac{\partial}{\partial x_{ij}} \sigma_r(X) = (-1)^r \frac{\partial}{\partial x_{ij}} x_{n-r},$$

which by Theorem 1 reduces to $(-1)^{r-1}(X_{n-r})_{ji} = (\Sigma_r)_{ji}$. ■

We are thus justified in calling Σ_r the "rth adjoint" of X, and we write

$$\Sigma_r = \operatorname{adj}_r(X). \tag{4.4}$$

In particular $\Sigma_0 = I = \Sigma_1$ and $\Sigma_n = \operatorname{adj}(X)$.

The "rth adjoint" may also be expressed as

$$\Sigma_r = \sum_{\boldsymbol{\alpha}} U_\alpha \operatorname{adj}(U_\alpha^T X U_\alpha) U_\alpha^T, \tag{4.5}$$

where $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_r)$ and $U_\alpha = [\mathbf{e}_{\alpha_1},\dots,\mathbf{e}_{\alpha_r}]$. Indeed, this follows from the two facts that

$$\mathbf{e}_j^T U_\alpha = \begin{cases} \mathbf{e}_q^T & \text{if } j = \alpha_q, \\ 0 & \text{if } j \notin \{\alpha_1,\dots,\alpha_r\} \end{cases}$$

and that if $i = \alpha_p$, $j = \alpha_q$, then $(\partial/\partial x_{ij})|X^{\boldsymbol{\alpha}}_{\boldsymbol{\alpha}}| = (\operatorname{adj}(X^{\boldsymbol{\alpha}}_{\boldsymbol{\alpha}}))_{(p,q)}$. We may in fact go one step further, and expand $\operatorname{adj}(\mu I - (\lambda I - X))$ to show that

$$\operatorname{adj}_r(X - \lambda I) = \sum_{t=0}^{r-1} (-1)^t \binom{n-r+t}{n-r} \operatorname{adj}_{r-t}(X)\lambda^t. \tag{4.6}$$

A second application of Theorem 2 is the fact that

COROLLARY 2.

$$\tilde{X}^T = -\frac{X_k^T}{x_k} = \frac{\nabla_X x_k}{x_k} = \nabla_X \ln x_k. \tag{4.7}$$

This says that the potential function of the g-inverse of $\tilde{X}^T$ is $[\ln x_k]$, which is similar to the case of a two dimensional Green's function.

It should be remarked here that there are two "limiting" processes that have to be carried out, namely $\nabla_X(\cdot)$ and $\lim_{X\to A}$. Moreover, these limits cannot be interchanged. For example, we are not allowed to conclude that $[\nabla_X\Delta(X,\lambda)]_{X=A}=\nabla_X(\lambda^k\Delta(U,\lambda))]_{X=A}=\lambda^k[\nabla_X(\Delta(U,\lambda))]_{X=A}$. In fact if A is a matrix with $k(A)=k_0$, $l(A)=l_0$, then $\nabla_X\Delta(X,\lambda)]_{X=A}=-(A^T_{k_0-l_0}\lambda^{k_0-l_0}+\cdots+A^T_{n-1}\lambda^{n-1})=-\operatorname{adj}(\lambda I-A)^T$, while $\lambda^k[\nabla_X\Delta(U,\lambda)]_{X=A}$ becomes $-\lambda^{k_0}[A^T_{k_0}+A^T_{k_0+1}\lambda+\cdots+A^T_{n-1}\lambda^{n-k_0}]$.

Combining Theorems 1 and 2, we may now obtain the potential function for X^d.

THEOREM 3. *Let $A\in\mathbb{F}_{n\times n}$ have zero order $k(A)=k_0$ and index $l(A)=l_0$. Let $a_i=x_i(A)$, $A_i=X_i(A)$, and $\alpha_i=\chi_i(A)$ denote the coefficients of λ^i in the expansions of $|\lambda I-A|$, $\operatorname{adj}(\lambda I-A)$, and $\Delta(U_A,\lambda)^{-1}$ respectively. Also let*

$$g_{k,t}=\sum_{i=0}^{t(X)}\chi_i x_{k-i}\quad \textit{and}\quad f_{k,t}=x_k g_{k,t},$$

where $t=t(X)$ is an integer valued function of X such that $l(X)\leqslant t(X)\leqslant k(X)$ for all $X\in\mathbb{F}_{n\times n}$. Then

$$(A^d)^T=[\nabla_X\ln f_{k,t}]_{X=A}. \tag{4.8}$$

Proof. First of all, $\nabla\ln(x_k g_{k,t})=\nabla\ln x_k+\nabla\ln g_{k,t}$ in which $\nabla\ln x_k=(1/x_k)\nabla x_k=-X_k^T/x_k=\tilde{X}$. Next consider $\nabla\ln g_{k,t}=\Sigma_{i=0}^t(\nabla\chi_i)x_{k-i}+\Sigma_{i=0}^t\chi_i(\nabla x_{k-i})$. In the first sum we separate off $(\nabla\chi_0)x_k=\nabla(1/x_k)\cdot x_k=-\nabla x_k/x_k=-\tilde{X}$. This approaches $-\tilde{A}$ as $X\to A$. The remaining terms $\Sigma_{i=1}^t(\nabla\chi_i)x_{k-i}$ will all vanish as $X\to A$, $k\to k_0$, and $l\to l_0$, since $a_{k_0-i}=0$ for all $i\geqslant 1$. The second sum reduces with the aid of Theorem 2 to $[-\chi_0X_k^T-\Sigma_{i=1}^t\chi_iX_{k-i}^T]_{X=A}=\tilde{A}-\Sigma_{i=1}^{l_0}\alpha_iA^T_{k_0-i}$, because $A_{k_0-i}=0$ for $i>l_0$. Collecting terms yields $\nabla\ln x_kg_{k,t}=\tilde{A}-\Sigma_0^{l_0}\alpha_iA^T_{k_0-i}$, which by (3.10) reduces to $(A^d)^T$, completing the proof. ∎

REMARKS.

(1) The most convenient choices for the potential function $f_{k,t}=x_kg_{k,t}$ are obtained by selecting $t(X)\equiv l(X)$ or $t(X)\equiv k(X)$, giving for example

$$f_{k,l}=x_k(\chi_0x_k+\chi_1x_{k-1}+\cdots+\chi_lx_{k-l}). \tag{4.9}$$

Alternatively we could for example take $t(X)=[\{l(X)+k(X)\}/2]$, even though this looks a little artificial.

(2) By analogy to (1.2) we may rewrite the potential function for $(A^d)^T$ as the logarithm of a suitable determinant. We shall need the following auxiliary result.

LEMMA 1. *If* $M\mathbf{y}=\mathbf{e}_1$, *where*

$$M=\begin{bmatrix}\mathbf{m}_1^T\\ M_2\end{bmatrix}$$

is $n\times n$ *and invertible, then*

$$\mathbf{b}^T\mathbf{y}=\frac{1}{|M|}\det\begin{bmatrix}\mathbf{b}^T\\ M_2\end{bmatrix}. \tag{4.10}$$

Proof. Use Cramer's rule. ∎

We may apply this to the case where

$$M_t=\begin{bmatrix}x_k & & & \bigcirc\\ x_{k+1} & x_k & & \\ \vdots & & \ddots & \\ x_{k+t} & & \cdots & x_k\end{bmatrix},\qquad \mathbf{y}=\begin{bmatrix}\chi_0\\ \chi_1\\ \vdots\\ \chi_t\end{bmatrix},\quad\text{and}\quad \mathbf{b}=\begin{bmatrix}x_k\\ x_{k-1}\\ \vdots\\ x_{k-t}\end{bmatrix}.$$

Then by Lemma 1,

$$g_{k,t}=\chi_0x_k+\cdots+\chi_tx_{k-t}=\mathbf{b}^T\mathbf{y}$$

$$=\frac{1}{x_k^{t+1}}\left|\begin{array}{c|cccccc}x_k & x_{k-1} & \cdot & \cdot & \cdot & x_{k-t} & \\ \hline x_{k+1} & x_k & & & & \bigcirc & \\ \cdot & x_{k+1} & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & & \cdot & \cdot & & \\ \cdot & \cdot & & & \cdot & \cdot & \\ x_{k+t} & x_{k+t-1} & \cdot & \cdot & \cdot & x_{k+1} & x_k\end{array}\right|.$$

Hence the potential function $f_{k,t}$ can be written as $|W|$, where

$$W = W_{k,t} = \frac{1}{x_k}\left[\begin{array}{c|cccccc} x_k^2 & x_k x_{k-1} & \cdot & \cdot & \cdot & \cdot & x_k x_{k-t} \\ \hline x_{k+1} & x_k & & & & & \\ \cdot & x_{k+1} & \cdot & & & \bigcirc & \\ \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & & \cdot & \cdot & & \\ \cdot & \cdot & & & \cdot & \cdot & \\ x_{k+t} & x_{k+t-1} & \cdot & \cdot & \cdot & x_{k+1} & x_k \end{array}\right]_{(t+1)\times(t+1)} \tag{4.11}$$

For $t = k$, this yields (2.3).

5. FURTHER REPRESENTATIONS FOR A^d

The expression for the Drazin inverse in (3.10) has the form of a convolution. Indeed it is easily seen that the sum $\sum_{i=0}^{l}\chi_i X_{k-i}$ equals the coefficient of λ^k in

$$\begin{aligned} G(\lambda) &= \operatorname{adj}(\lambda I - X)\Delta^{-}(U,\lambda) \\ &= (\lambda^{k-l}X_{k-l} + \cdots + \lambda^{n-1}I)(\chi_0 + \chi_1\lambda + \cdots + \chi_l\lambda^l)I. \end{aligned} \tag{5.1}$$

It is instructive to derive this in a more direct way, which not only allows us to compute the potential functions for $E = I - XX^d$ and $N = X(I - XX^d)$, but also allows us to draw a close parallel with the complex matrix case.

Using the Fitting representation (2.2), we have

$$G(\lambda) = Q\begin{bmatrix} \lambda^k \operatorname{adj}(\lambda I - U) & 0 \\ 0 & \Delta(U,\lambda)\operatorname{adj}(\lambda I - \eta) \end{bmatrix} Q^{-1}\cdot\Delta^{-}(U,\lambda). \tag{5.2}$$

Now $\operatorname{adj}(\lambda I - U) = U_0 + U_1\lambda + \cdots + U_{n-k-1}\lambda^{n-k-1}$, where $\chi_0 U_0 = -U^{-1}$, while $\operatorname{adj}(\lambda I - \eta) = (\eta^{k-1} + \eta^{k-2}\lambda + \cdots + I\lambda^{k-1})$. Since $\Delta(U,\lambda)\Delta^{-}(U,\lambda) = 1 - \lambda^{l+1}s(\lambda)$, we see that the coefficient of λ^k in $G(\lambda)$ is precisely

$$Q\begin{bmatrix} \chi_0 U_0 & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} = -X^d.$$

Similarly the coefficients of λ^{k-1} and λ^{k-2} become $E = I - XX^d$ and $N = X(I - XX^d)$ respectively.

Using the extended definition of the operator $(1/r!)\mathbb{D}_\lambda^r[\cdot]_{\lambda=0}$, we may replace $\Delta^-(U,\lambda)$ by $1/\Delta(U,\lambda)$ in (5.1). This gives the following:

THEOREM 3. *Let $X \in \mathbb{F}_{n\times n}$ and suppose its characteristic and adjoint polynomials are given as in* (2.1) *and* (2.5). *Further let* $G(\lambda) = \operatorname{adj}(\lambda I - X)/\Delta(U,\lambda)$. *Then*

$$-X^d = \frac{1}{k!}\mathbb{D}_\lambda^k[G(\lambda)]_{\lambda=0}, \tag{5.3a}$$

$$E = I - XX^d = \frac{1}{(k-1)!}\mathbb{D}_\lambda^{k-1}[G(\lambda)]_{\lambda=0}, \tag{5.3b}$$

$$N = X(I - XX^d) = \frac{1}{(k-2)!}\mathbb{D}_\lambda^{k-2}[G(\lambda)]_{\lambda=0}, \tag{5.3c}$$

where $k = k(X)$ is the zero order of X.

In particular if $k_0 = k(A)$, $l_0 = l(A)$, then

$$\begin{aligned} E = I - AA^d &= \sum_{i=0}^{l_0-1} \alpha_i A_{k_0-i-1}, \\ N = A(I - AA^d) &= \sum_{i=0}^{l_0-2} \alpha_i A_{k_0-i-2}, \end{aligned} \tag{5.4}$$

which may also be obtained from the spectral theorem [4].

Analogous representations can be given for all members of the Drazin chain [6]

$$(N^{l-1}, N^{l-2}, \ldots, N, E, -X^d, -X^{d^2}, \ldots). \tag{5.5}$$

This comes as no surprise if one recalls the complex Laurent expansion [7, Theorem 2.1] for the resolvent:

$$(\lambda I - X)^{-1} = \frac{N^{l-1}}{\lambda^l} + \frac{N^{l-2}}{\lambda^{l-2}} + \cdots + \frac{E}{\lambda} - X^d - X^{d^2}\lambda - \cdots, \tag{5.6}$$

where $N=(I-XX^d)X$, $E=I-XX^d$, and λ is small. Hence

$$G(\lambda)=\frac{\operatorname{adj}(\lambda I-X)}{\Delta(U,\lambda)}=N^{l-1}\lambda^{k-l}+\cdots+N\lambda^{k-2}+E\lambda^{k-1}$$
$$-X^d\lambda^k-X^{d^2}\lambda^{k+1}+\cdots,$$

from which the coefficients of λ^i may be extracted by differentiation, which is precisely what (5.3) does formally. For complex matrices, the representation (5.3a) can also be obtained from the contour integral representation for X^d as given in [5]:

$$X^d=\frac{1}{2\pi i}\int_\Gamma(\lambda I-X)^{-1}\frac{d\lambda}{\lambda}=\frac{1}{2\pi i}\int_\Gamma G(\lambda)\frac{d\lambda}{\lambda^{k+1}},\tag{5.7}$$

where the contour Γ encloses the spectrum of X except for the origin. Using the residue theorem, this equals $-\operatorname{Res}_{\lambda=0}[G(\lambda)/\lambda^k]$, which by the differentiation formula for the residue at a pole of order k yields the desired result.

A little-known fact is that the X^d can actually be calculated from a knowledge of $E=I-XX^d$ and the g-inverse $\tilde{X}=-X_k/x_k$. Indeed, from (3.5) we obtain

$$X^d=X^l(\tilde{X})^{l+1}=(X^l\tilde{X}^l)\tilde{X}=(I-E)\tilde{X},\tag{5.8}$$

which in turn will yield (3.10) if we substitute the expression (5.4) for E.

Let us conclude with several pertinent remarks.

REMARKS.

(1) The potential functions for each member of the Drazin chain (5.5) can be computed as in Theorem 3. For example, it is easily verified that

$$\begin{aligned}[\nabla_X g_{k-2,t}]_{X=A}&=-N_A^T,\\ [\nabla_X g_{k-1,t}]_{X=A}&=-E_A^T,\end{aligned}\tag{5.9}$$

where $l_0\leqslant t(A)\leqslant k_0$.

(2) For small values of l, the potential function $f_{k,l}=x_k g_{k,l}$ of X^d can be expressed as a small determinant. For example, if $l=0$, then $k=0$, $g_{k,l}=1$ and $-(A^d)^T=\nabla_X \ln x_0$, which becomes (1.2). On the other hand, if $l=1$, then

$$f_{k,1}=\frac{1}{x_k}\begin{vmatrix}x_k & x_{k-1}\\ x_{k+1} & x_k\end{vmatrix}=\frac{x_k^2-x_{k-1}\cdot x_k}{x_k}.$$

Lastly, if $l = 2$, then

$$f_{k,2} = \frac{1}{x_k^2}\begin{vmatrix} x_k & x_{k-1} & x_{k-2} \\ x_{k+1} & x_k & 0 \\ x_{k-2} & x_{k+1} & x_k \end{vmatrix}.$$

(3) Using (2.4), the potential function $f_{k,t}$ can dually be expressed in terms of σ_r and the core rank $\rho = n - k$. This gives

$$f_{k,t} = \frac{(-1)^{\rho}}{\sigma_\rho t}\left|\begin{array}{c|cccc} \sigma_\rho & -\sigma_{\rho+1} & \cdots & (-1)^t\sigma_{\rho+t} \\ \hline -\sigma_{\rho-1} & \sigma_\rho & & \\ \vdots & -\sigma_{\rho-1} & \ddots & \bigcirc \\ (-1)^t\sigma_{\rho-t} & \cdots & -\sigma_{\rho-1} & \sigma_\rho \end{array}\right|.$$

(4) In [3] the same problem is considered for the Moore-Penrose generalized inverse.

REFERENCES

1. A. Ben Israel and T. N. E. Greville, *Generalised Inverses, Theory and Applications*, Wiley, New York, 1974.
2. R. Gabriel, Das verallgemeinerte Inverse einem Matrix in einem beliebigen Korper analytisch betrachtet, *J. Reine Ange.* Math. 244:83–93 (1970).
3. ______, Das allgemeine element der verallgemeinerten inversen von Moore und Penrose, Revue Rounaine de Math. Pures et Appl. XXVII 6:689–702 (1982).
4. R. E. Hartwig, More on the Souriau-Frame Algorithm and the Drazin inverse, *SIAM J. Appl. Math.* 31:42–46 (1976).
5. ______, "On the derivative of the Drazin inverse of a complex matrix, *SIAM J. Math. Anal.* 10:207–216 (1979).
6. ______, Theory and applications of matrix perturbations, with respect to Hankel matrices and power formulae, *Internat. J. System Sci.* 10:437–460 (1979).
7. N. J. Rose, The Laurent Expansion of a generalized resolvent, with some applications, *SIAM J. Math. Anal.* 9:751–758 (1978).

Received 22 September 1983; revised 15 November 1983