ELSEVIER

Mathematics and Computers in Simulation 42 (1996) 585–593

MATHEMATICS AND COMPUTERS IN SIMULATION

Various proofs of Sylvester's (determinant) identity

Alkiviadis G. Akritas [a,*], Evgenia K. Akritas [a], Genadii I. Malaschonok [b]

[a] University of Kansas, Department of Computer Science, Lawrence, KS 66045-2192, USA

[b] Kiev University, Department of Cybernetics, Vladimirska 64, U-252017 Kiev, Ukraine

Abstract

Despite the fact that the importance of Sylvester's determinant identity has been recognized in the past, we were able to find only one proof of it in English (Bareiss, 1968), with reference to some others. (Recall that Sylvester (1857) stated this theorem without proof.) Having used this identity, recently, in the validity proof of our new, improved, matrix-triangularization subresultant polynomial remainder sequence method (Akritas et al., 1995), we decided to collect all the proofs we found of this identity – one in English, four in German and two in Russian, in that order – in a single paper (Akritas et al., 1992). It turns out that the proof in English is identical to an earlier one in German. Due to space limitations two proofs are omitted.

Keywords: Sylvester's identity; Polynomial remainder sequence (prs); Matrix-triangularization subresultant prs

1. Introduction

Throughout this paper we consider an $n \times n$ matrix $A = (a_{ij})$ $(i, j = 1, 2, \ldots, n)$ with elements a_{ij} and determinant $|A|$, also written det A. We introduce the notation

$$a_{ij}^{k} = \begin{vmatrix} a_{11} & \cdots & a_{1,k-1} & a_{1j} \\ \vdots & \ddots & \vdots & \vdots \\ a_{k-1,1} & \cdots & a_{k-1,k-1} & a_{k-1,j} \\ a_{i1} & \cdots & a_{i,k-1} & a_{ij} \end{vmatrix},$$

with $1 \leqslant k \leqslant n$, and $k \leqslant i, j \leqslant n$; that is, this determinant, of order k, is obtained from the matrix A by adding row i and column j to the upper left corner minor of order $k-1$.

If we set $a_{00}^{0} = 1$, then for $1 \leqslant p \leqslant n$, Sylvester's identity can be expressed as

$$\det D_p(A) = (\det A) \cdot (a_{p-1,p-1}^{p-1})^{n-p}, \tag{1}$$

where

* Corresponding author. Partially supported by GRF grant 3089-XX-0038 (1993) of the University of Kansas.

PII S0378-4754(96)00035-3

$$D_p(A) = \begin{pmatrix} a_{pp}^p & a_{p,p+1}^p & \cdots & a_{pn}^p \\ a_{p+1,p}^p & a_{p+1,p+1}^p & \cdots & a_{p+1,n}^p \\ \vdots & \vdots & \ddots & \vdots \\ a_{np}^p & a_{n,p+1}^p & \cdots & a_{nn}^p \end{pmatrix}$$

of order $n-p+1$ and $a_{ij}^p (p, i, j = 1, 2, \ldots, n)$ are minors, of order p, obtained from matrix A by adding row i and column j to the (upper left) corner minor of order $p-1$ (see for example [3, 6, 7, 8, pp. 30–35]). Note that

$$(a_{ij}^1) = (a_{ij}) \qquad (i, j = 1, 2, \ldots, n),$$

as well as the fact (easily shown using any of the proofs of the following section) that (1) is valid over an integral domain.

To see an application of Sylvester's identity consider the process of matrix-triangularization over the integers [7]; it is assumed that all diagonal minors of A are different from zero:

$$a_{kk}^k \neq 0, \quad k = 1, 2, \ldots, n.$$

In the first step of this process, matrix $A(= A^{(1)})$ is transformed in the following way: from each row with index i, $i \geqslant 2$, multiplied times a_{11}, subtract the first row multiplied times a_{i1}. The elements of the resulting matrix $A^{(2)}$ are as follows:

$$a_{1j}^2 = a_{1j}^1, \quad a_{i1}^2 = 0 \quad \text{and} \quad a_{ij}^2 \quad \text{for } i > 1,$$

where the notation was introduced above. Observe that the elements of $A^{(2)}$ have grown in size and the same is expected of the elements of each subsequently updated matrix. This growth needs to be controlled in order to avoid wasting computer memory and time. Using (1), at the kth step, $k \geqslant 2$, the elements of $A^{(k+1)}$ can be reduced by dividing out a diagonal minor of order $k-1$. To see this, assume that at the $(k-1)$th step we obtained matrix $A^{(k)}$ with elements

$$a_{ij}^k = \begin{cases} a_{ij}^i & \text{for } i = 1, 2, \ldots, k-1, \quad j \geqslant i, \\ 0 & \text{for } i > j, \quad j = 1, 2, \ldots, k-1, \\ a_{ij}^k & \text{for } i, j \geqslant k. \end{cases}$$

Then, at the kth step matrix $A^{(k)}$ is transformed as follows: from each row with index i, $i \geqslant k+1$, multiplied times a_{kk}^k, subtract the row with index k, multiplied times a_{ik}^k; then the (i, j)th element of row i, $j \geqslant k+1$, is of the form

$$a_{kk}^k \cdot a_{ij}^k - a_{ik}^k \cdot a_{kj}^k, \tag{2}$$

and in the first k columns there are zeros. Consider now (1) with $\det A = a_{ij}^{k+1}$, of order $k+1$, for $k < i \leqslant n$, $k < j \leqslant n$, and under the condition $p = k$; we then obtain

$$\begin{vmatrix} a_{kk}^k & a_{kj}^k \\ a_{ik}^k & a_{ij}^k \end{vmatrix} = a_{ij}^{k+1} \cdot a_{k-1,k-1}^{k-1}$$

from which we see that (2), and hence all the updated elements of $A^{(k+1)}$, can be reduced by dividing out the diagonal minor of order $k-1$.

2. The various proofs

In this section we present all seven proofs of Sylvester's identity (1). However, due to space restrictions, only three are presented in full: the one by Bareiss, one proved with the help of Jacobi's Theorem and one by Malaschonok; a brief outline is given for the rest. The whole presentation is in the alphabetical order of the language they were written in; namely, we first present Bareiss's proof (written in English), then the four proofs found in Kowalewski's book (written in German) and finally Malaschonok's proofs (written in Russian). We will also denote them as B, K1, K2, K3, K4, M1 and M2, respectively, and we will try to compare them.

2.1. Bareiss's proof (B)

This proof of (1) [3] is a compact form of K2, the proof found in Kowalewski's book [6, pp. 91–93], and is based on the fact that

$$A^{-1} = \frac{1}{|A|} \cdot (\text{adj } A), \tag{3}$$

where the determinant of A is assumed $\neq 0$, and (adj A) is the $n \times n$ matrix whose (i, j)th entry $(\text{adj } A)_{ij}$ is $(-1)^{i+j}$ det $A_{ji} = \alpha_{ij}$; α_{ji} is the *algebraic complement* of a_{ij} and A_{ij} is the matrix obtained by crossing out the ith row and the jth column:

$$(\text{adj } A) = (\alpha_{ij})^{\mathrm{T}},$$

the transpose of (α_{ij}), where $\alpha_{ij} = (-1)^{i+j}$ det A_{ij}. The above equation is a direct corollary of the well known fact that

$$(\text{adj } A) \cdot A = |A| \cdot I,$$

where I is the identity matrix. (See also Sections 2.3 and 2.7.)

To start the proof of (1), we partition A and factor by block triangularization such that

$$A = \begin{pmatrix} W & X \\ Y & Z \end{pmatrix} = \begin{pmatrix} W & 0 \\ Y & I \end{pmatrix} \cdot \begin{pmatrix} I & W^{-1}X \\ 0 & Z - YW^{-1}X \end{pmatrix},$$

where W is a nonsingular square matrix of order $p-1$. Then

$$|A| = |W| \cdot |Z - YW^{-1}X|. \tag{4}$$

Multiplying both sides by $|W|^{n-p}$, Eq. (4) becomes

$$|W|^{n-p}|A| = ||W| \cdot |Z - YW^{-1}X|| \tag{5}$$

because the second determinant on the right-hand side of (4) is of order $(n - p + 1)$. Consider a_{ij}^p, the determinant of order p $(p \leqslant i, j \leqslant n)$ which is $|W|$ bordered by a row and a column. If we apply now the identity (4) to each element of the determinant on the right-hand side of (5) we obtain

$$|W| \cdot \left(a_{ij} - \sum_{r=1}^{p-1} \sum_{s=1}^{p-1} a_{ir}(W^{-1})_{rs} a_{sj} \right) = a_{ij}^p \quad (p \leqslant i, j \leqslant n). \tag{6}$$

Since $|W| = a^{p-1}_{p-1,p-1}$, Eq. (5) takes the form of Sylvester's identity (1).

The validity of Eq. (6) can be seen if we expand a^p_{ij} by minors according to the last column, then expand each minor (except $|W|$) according to the last row and, finally, make use of (3).

2.2. *First proof in Kowalewski's book (K1)*

This proof of (1), due to Studnička [6, pp. 76–80], is based on induction. It is straightforward and rather long, no special theorems are used, and can be understood even by high school students. For details see [2] or [6].

2.3. *Second proof in Kowalewski's book (K2)*

This proof of (1) [6, pp. 91–93] is the long version of the one presented by Bareiss. Namely, both B and K2 are different versions of one and the same proof, since they are based on the identity

$$a^p_{rs} = a_{rs}a^{p-1}_{p-1,p-1} - \sum_{\rho=1}^{p-1}\sum_{\sigma=1}^{p-1} a_{\rho s}a_{r\sigma}\overline{a_{\rho\sigma}},$$

where $\overline{a_{\rho\sigma}}$ denotes the algebraic complement of $a_{\rho\sigma}$ in $a^{p-1}_{p-1,p-1}$.

The only difference between the two proofs is that, in K2, the composition of rows, premultiplied times a coefficient, corresponds to matrix multiplication in B. (See also Sections 2.1 and 2.7.)

2.4. *Third proof in Kowalewski's book (K3)*

This is an elegant proof of (1) [6, pp. 93–94] based on Jacobi's identity [5], which can be stated as follows:

Jacobi (1841): *Let* $|A| \neq 0$ *be a nonvanishing determinant, and let* $|(\text{adj } A)|$ *be its adjoint determinant. Moreover, let* $|(\text{adj}\tilde{A})^p|$, *of order p, be a minor of* $|(\text{adj } A)|$, *and let* $\tilde{a}^p_{rs}$ *be the corresponding minor of* $|A|$. *Then* $|(\text{adj } \tilde{A})^p|$ *differs from the algebraic complement of* $\tilde{a}^p_{rs}$ *by the factor* $|A|^{p-1}$.

The third proof of (1) can now be stated as follows. Consider $|A|$ as defined above, and its adjoint determinant $|(\text{adj } A)|$, defined as

$$|(\text{adj } A)| = \begin{vmatrix} (\text{adj } A)_{11} & (\text{adj } A)_{12} & \dots & (\text{adj } A)_{1n} \\ (\text{adj } A)_{21} & (\text{adj } A)_{22} & \dots & (\text{adj } A)_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (\text{adj } A)_{n1} & (\text{adj } A)_{n2} & \dots & (\text{adj } A)_{nn} \end{vmatrix}, \tag{7}$$

where we follow Kowalewski's notation now and do not transpose the matrix. In (7) consider the minor

$$|(\text{adj } \tilde{A})^{n-p+1}| = \begin{vmatrix} (\text{adj } A)_{pp} & \dots & (\text{adj } A)_{pn} \\ \vdots & \ddots & \vdots \\ (\text{adj } A)_{np} & \dots & (\text{adj } A)_{nn} \end{vmatrix}. \tag{8}$$

If we now delete in (8) row r and column s $(r, s = p, \dots, n)$, we obtain a determinant of order $(n-p)$, which by Jacobi's Theorem above is equal to

$$|A|^{n-p+1} \cdot \begin{vmatrix} a_{11} & \dots & a_{1,p-1} & a_{1s} \\ \vdots & \ddots & \vdots & \vdots \\ a_{p-1,1} & \dots & a_{p-1,p-1} & a_{p-1,s} \\ a_{r1} & \dots & a_{r,p-1} & a_{rs} \end{vmatrix} = |A|^{n-p+1} a_{rs}^{p}$$

multiplied times $(-1)^{r+s}$.

However, the algebraic complement of $(\text{adj } A)_{rs}$ in $|(\text{adj } \tilde{A})^{n-p+1}|$ is also

$$|A|^{n-p+1} a_{rs}^{p}$$

and, therefore, the determinant adjoint to $|(\text{adj } \tilde{A})^{n-p+1}|$ has the value

$$|A|^{(n-p+1)(n-p-1)} \cdot \begin{vmatrix} a_{pp}^{p} & \dots & a_{pn}^{p} \\ \vdots & \ddots & \vdots \\ a_{np}^{p} & \dots & a_{nn}^{p} \end{vmatrix}.$$

On the other hand, since $|(\text{adj } A)| = |A|^{n-1}$, this adjoint determinant is equal to $|(\text{adj } \tilde{A})^{n-p+1}|^{n-p}$ or, since (by Jacobi's Theorem)

$$|(\text{adj } \tilde{A})^{n-p+1}| = \begin{vmatrix} a_{11} & \dots & a_{1,p-1} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \dots & a_{p-1,p-1} \end{vmatrix} \cdot |A|^{n-p},$$

this adjoint determinant is equal to

$$|A|^{(n-p)^2} \cdot \begin{vmatrix} a_{11} & \dots & a_{1,p-1} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \dots & a_{p-1,p-1} \end{vmatrix}^{n-p}.$$

From the above it follows that

$$\begin{vmatrix} a_{pp}^{p} & \dots & a_{pn}^{p} \\ \vdots & \ddots & \vdots \\ a_{np}^{p} & \dots & a_{nn}^{p} \end{vmatrix} = |A| \cdot \begin{vmatrix} a_{11} & \dots & a_{1,p-1} \\ \vdots & \ddots & \vdots \\ a_{p-1,1} & \dots & a_{p-1,p-1} \end{vmatrix}^{n-p},$$

which is Sylvester's identity.

2.5. Fourth proof in Kowalewski's book (K4)

This gives us more than just the proof of Sylvester's identity (1); namely, it proves [6, pp. 100–101] the generalized identity of Sylvester, also known as the Sylvester–Kowalewski identity, a partial instance of which is Sylvester's identity.

The generalized identity can be expressed as

$$\det\ D_m(A) = (a^{p-1}_{p-1,p-1})^{\binom{n-p}{m-p+1}} \cdot |A|^{\binom{n-p}{m-p}}, \tag{9}$$

from which we see that, for $m = p$, (4) reduces to (1).

2.6. *Malaschonok's first proof (M1)*

This proof of (1) [7] differs from the other ones in that it does *not* require matrix A to be defined over a field, and, therefore, does not require a special proof for the case $\det\ A = 0$, based on the continuity of the determinants. Hence, M1 applies to matrices over any commutative ring.

This proof is based on the following equation (which is obtained by expanding a^p_{ij} according to the last column):

$$a^p_{ij} = a^{p-1}_{p-1,p-1} a_{ij} - \sum_{s=1}^{p-1} a_{sj} a^{p-1}_{(si)}, \tag{10}$$

where $2 \leqslant p \leqslant n$, $1 \leqslant i, j \leqslant n$, $a^{p-1}_{(si)}$ is the minor of order $p-1$, which is located in the upper left corner of A, if in A we replace row s $(s \leqslant p)$ by row i, and

$$a^p_{ij} = 0 \quad \text{for } 2 \leqslant p \leqslant n,\ \ i < p,\ \ j \leqslant n$$
$$\text{or for } 2 \leqslant p \leqslant n,\ \ i \leqslant n\ \ j < p. \tag{11}$$

To prove Sylvester's identity we transform matrix A in the following way: we multiply each row with number i, $p \leqslant i \leqslant n$, times $a^{p-1}_{p-1,p-1}$ and subtract from it each one of the rows with numbers s, $1 \leqslant s < p$, multiplied times the minor $a^{p-1}_{(si)}$. The resulting matrix is then

$$T = (t_{ij}), \quad i, j = 1, 2, \ldots, n,$$

with elements

$$t_{ij} = \begin{cases} a_{ij}, & 1 \leqslant i < p+1,\ \ 1 \leqslant j \leqslant n, \\ a^{p-1}_{p-1,p-1} a_{ij} - \sum_{s=1}^{p-1} a_{sj} a^{p-1}_{(si)}, & p+1 \leqslant i \leqslant n,\ \ 1 \leqslant j \leqslant n, \end{cases}$$

and determinant

$$\det\ T = (\det\ A) \cdot (a^{p-1}_{p-1,p-1})^{n-p}. \tag{12}$$

If we make use of (10) and (11), then it is possible to write the matrix T as follows :

$$T = \begin{pmatrix} a_{11} & \cdots & a_{1,p-1} & a_{1p} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,p-1} & a_{2p} & \cdots & a_{2n} \\ \vdots & \ddots & & & & \vdots \\ a_{p1} & \cdots & a_{p,p-1} & a_{pp} & \cdots & a_{pn} \\ 0 & \cdots & 0 & a^{p}_{p+1,p} & \cdots & a^{p}_{p+1,n} \\ \vdots & & \vdots & & & \vdots \\ 0 & \cdots & 0 & a^{p}_{np} & \cdots & a^{p}_{nn} \end{pmatrix}.$$

If we expand the above matrix T according to the minors contained in the first p rows, using Laplace's theorem

$$|A| = \sum |M||\overline{M}|,$$

where $\overline{M}$ is the algebraic complement of M, and the sum is taken over the $\binom{n}{m}$ products (m being the specified rows containing the minors), we have

$$\det T = \sum_{s=p}^{n} (-1)^{p+s} \left| T \begin{bmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p-1, s \end{bmatrix} \right| \cdot \left| \overline{T} \begin{bmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p-1, s \end{bmatrix} \right|, \tag{13}$$

where in the brackets above, the first row indicates which rows of T are used and the second row indicates which columns of T are used.

Here we consider that all minors of the type

$$\left| \overline{T} \begin{bmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p-1, s \end{bmatrix} \right|,$$

which are obtained by crossing out rows $1, 2, \dots, p$ and columns $q_1, q_2, \dots, q_p$, are equal to zero, except for those minors in which among the crossed out columns $q_1, q_2, \dots, q_p$ are contained *all* the zero columns $1, 2, \dots, p-1$.

Comparing the matrices A, $D_p(A)$ and T, we observe that there exist the following two equalities of minors for $s \geqslant p$:

$$\left| T \begin{bmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p-1, s \end{bmatrix} \right| = \left| A \begin{bmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p-1, s \end{bmatrix} \right| = a^{p}_{ps},$$

$$\left| \overline{T} \begin{bmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p-1, s \end{bmatrix} \right| = \left| \overline{D_p(A)} \begin{bmatrix} p \\ s \end{bmatrix} \right|.$$

Therefore, Eq. (13) now becomes

$$\det T = \sum_{s=p}^{n} (-1)^{1+(s-p+1)} a^{p}_{ps} \left| \overline{D_p(A)} \begin{bmatrix} p \\ s \end{bmatrix} \right|.$$

However, the right-hand side of the above equation represents the expansion of the determinant of the matrix $D_p(A)$ by the first row. Therefore, we arrive at the equality

$$\det D_p(A) = \det T. \tag{14}$$

Finally, from Eqs. (12) and (14) immediately follows Sylvester's identity (1).

2.7. Malaschonok's second proof (M2)

This proof is based on (10) and (11) of Section 2.6.
To begin the proof, let

$$A = \begin{pmatrix} W & X \\ Y & Z \end{pmatrix},$$

where W is a square submatrix of order $p-1$; more precisely, we have $\det W = a^{p-1}_{p-1,p-1}$. Moreover, let

$$I_{p-1} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

be the identity diagonal matrix of order $p-1$, and let

$$I_{n-p+1}(a) = \begin{pmatrix} a^{p-1}_{p-1,p-1} & & 0 \\ & \ddots & \\ 0 & & a^{p-1}_{p-1,p-1} \end{pmatrix}$$

be the diagonal matrix of order $n-p+1$, with elements $a^{p-1}_{p-1,p-1}$.
Set

$$[i, j] = \det W_{(ij)},$$

where $W_{(ij)}$ is obtained from W by replacing row i by row j (i.e. using the convention established in Section 2.6 we have $[i, j] = a^{p-1}_{(ij)}$) and consider the $(n-p+1) \times (p-1)$ matrix

$$A' = \begin{pmatrix} [1, p] & \dots & [p-1, p] \\ \vdots & \ddots & \vdots \\ [1, n] & \dots & [p-1, n] \end{pmatrix}.$$

Then, using (10) and (11), we have the matrix identity

$$\begin{pmatrix} I_{p-1} & 0 \\ A' & I_{n-p+1}(a) \end{pmatrix} A = \begin{pmatrix} W & X \\ 0 & D_p(A) \end{pmatrix}, \tag{15}$$

from which we obtain the determinantal identity

$$(a^{p-1}_{p-1,p-1})^{n-p+1} \cdot (\det A) = (\det D_p(A)) \cdot a^{p-1}_{p-1,p-1}$$

and after cancellations Sylvester's identity (1).

Comparing M2 with B, or K2, we see that B (and K2) are based on the inverse matrix transformation of (15); namely, they are based on

$$A = \begin{pmatrix} I_{p-1} & 0 \\ A' & I_{n-p+1}(a) \end{pmatrix}^{-1} \begin{pmatrix} W & X \\ 0 & D_p(A) \end{pmatrix}.$$

3. Conclusion

We are, thus, done with the presentation of the proofs of Sylvester's determinant identity (1). It is rather unfortunate that, due to "its inclusion as merely one of a class of theorems" [9, p. 60], none of the proofs of this identity can be found in the English-language books on matrix theory [4].

Despite all these, using Sylvester's identity we were able to obtain elegant proofs for the validity of the new, improved, matrix-triangularization subresultant prs method [1]. Therefore, there can be no doubt that, (not only) at the time of its discovery, this identity was – as Sylvester himself styled it – "a remarkable theorem" [9, p. 60].

References

[1] A.G Akritas, E.K. Akritas and G.I. Malaschonok, Computation of polynomial remainder sequences in integral domains, Reliable Computing 1 (4) (1995) 375–381.
[2] A.G. Akritas, E.K. Akritas and G.I. Malaschonok, Various proofs of Sylvester's (determinant) identity, Technical Report, University of Kansas, 1992.
[3] E.H. Bareiss, Sylvester's identity and multistep integer preserving Gaussian elimination, Math. Comp. 22 (1968) 565–578.
[4] R.A. Horn and C.A. Johnson, Topics in Matrix Analysis (Cambridge University Press, New York, 1991).
[5] C.G.J. Jacobi, Ueber die Bildung und die Eigenschaften der Determinanten (De formatione et proprietatibus Determinantum 1841), edited by Stäkel.
[6] G. Kowalewski, Einfürung in die Determinantentheorie (Chelsea, New York, 1948).
[7] G.I. Malaschonok, Solution of a system of linear equations in an integral domain, USSR J. Comp. Math. and Math. Phys. 23 (1983) 1497–1500 (in Russian).
[8] G.I. Malaschonok, System of linear equations over a commutative ring, Academy of Sciences of Ukraine Lvov, 1986 (in Russian).
[9] T. Muir, The theory of Determinants (Dover, New York).
[10] J.J. Sylvester, On the relation between the minor determinants of linearly equivalent quadratic functions, Philosophical Magazine 1 (Fourth Series) (1851) 295–305.