Elem. Math. 69 (2014) 33 – 39
0013-6018/14/010033-7
DOI 10.4171/EM/243

Elemente der Mathematik

A proof of the main theorem on Bezoutians

Branko Ćurgus and Aad Dijksma

B. Ćurgus has a PhD in mathematics from the University of Sarajevo. He enjoys teaching and researching beautiful mathematics.

A. Dijksma (PhD 1971) taught mathematics at the University of Groningen from 1972 to 2005. His research concerns operator theory including spectral theory and Schur analysis.

With two polynomials f and g and $n = \max\{\deg f, \deg g\}$ we associate an $n \times n$ matrix B, called the Bezoutian, and a $2n \times 2n$ matrix R, called the resultant. Their defining relations are given by (6) and (5) below, respectively. In terms of the coefficients of f and g they are given by (8) and (4). In this note we give a simple and self-contained proof of the equalities

$$\dim \ker B = \dim \ker R = \deg \gcd(f, g), \tag{1}$$

where gcd stands for greatest common divisor. H.K. Wimmer in [8] attributes this result to Jacobi who in 1836 showed that the singularity of what we call the Bezoutian implies the existence of a common factor of f and g. More contemporary proofs of (1) can be found in the recent books [3, Theorems 21.10 and 21.11] by H. Dym and [5, Theorem 8.30] by P.A. Fuhrmann. In the Introduction to [3, Chapter 21] it is shown that

The Frenchman Étienne Bézout (1730–1783) taught mathematics at the Garde du Pavillon, the Garde de la Marine and the Corps d'Artillerie and wrote several textbooks used widely in Europe and the USA. The little time left for research he devoted mainly to solving systems of equations in several variables. He developed the "method of simplifying assumptions": when the general problem appears unsoluble consider first special problems by making assumptions. He was successful: a theorem in algebraic geometry, an identity in elementary number theory, an integral domain and a matrix now carry his name. Our note concerns the Bézoutian matrix so termed by the Copley medalists James Joseph Sylvester (in 1853) and Arthur Cayley (in 1857). The Bézoutian matrix is a square matrix associated with two polynomials whose nullity equals the number of their common zeros counting multiplicities. We give a self-contained new proof of this fact. In the English literature the accent aigu on the e is often omitted.

$\dim\ker B \geq \deg\gcd(f, g)$ by using the defining formula for B, differentiation and chains of vectors. That equality prevails is then proved by using these chains and the so-called Barnett identity: $B = H_f g(C_f)$, where H_f is the Hankel matrix for f defined below and C_f is the companion matrix of f. In [5] the matrix B is expressed in terms of a matrix representation of $g(S_f)$, where S_f is the shift operator in the space X_f of polynomials modulo f, relative to two suitably chosen bases in X_f. In view of [5, Corollary 8.29] this formula is closely related to the Barnett identity. In this note we do not resort to this identity. Our approach, we think, is more direct. Of course some of the formulas derived below also appear in [3, Chapter 21] and [5, Chapter 8]. Our proofs of these formulas are different. For a survey of results related to Bezoutians, see [1, Fact 4.8.6] in the encyclopedic book by D.S. Bernstein, and for applications of Bezoutians in numerical linear algebra and system theory, see for example [4] and [6], respectively.

1 Notation and basic notions

The vector space of all polynomials with coefficients in $\mathbb{C}$ and in the variable z is denoted by $\mathbb{C}[z]$. Its Cartesian square is denoted by $\mathbb{C}^2[z]$. For $n \in \mathbb{N}$, $\mathbb{C}[z]_{<n}$ denotes the subspace of $\mathbb{C}[z]$ of all polynomials of degree strictly less than n. This space has dimension n. Similarly, $\mathbb{C}^2[z]_{<n}$ denotes the Cartesian square of $\mathbb{C}[z]_{<n}$.

We use I to denote the identity matrix, Z the reverse identity and N the nilpotent Jordan block:

$$Z := \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \cdots & 0 \end{bmatrix}, \qquad N := \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

For a polynomial

$$f(z) = f_0 + f_1 z + \cdots + f_n z^n$$

in $\mathbb{C}[z]$ we define two $n \times n$ matrices, one Hankel and one Toeplitz, associated with f as follows:

$$H_f := \begin{bmatrix} f_1 & \cdots & f_n \\ \vdots & \iddots & \vdots \\ f_n & \cdots & 0 \end{bmatrix}, \qquad T_f := \begin{bmatrix} f_0 & \cdots & f_{n-1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_0 \end{bmatrix}.$$

Since the left-multiplication by Z reverses the rows, it turns a Hankel matrix into a Toeplitz and vice versa:

$$ZH_f := \begin{bmatrix} f_n & \cdots & 0 \\ \vdots & \ddots & \vdots \\ f_1 & \cdots & f_n \end{bmatrix}, \qquad ZT_f := \begin{bmatrix} 0 & \cdots & f_0 \\ \vdots & \iddots & \vdots \\ f_0 & \cdots & f_{n-1} \end{bmatrix}.$$

As each Hankel matrix is symmetric, we have $ZT_f = \left(ZT_f\right)^\top = T_f^\top Z$, where the superscript $^\top$ is used to denote a matrix transpose. Consequently,

$$T_f^\top = ZT_f Z. \tag{2}$$

The vector space of upper (lower) triangular Toeplitz matrices is spanned by the identity I and the powers of N ($N^\top$, respectively). Therefore, the upper (lower) triangular Toeplitz matrices form a commutative algebra. In particular for polynomials f as above and $g(z) = g_0 + g_1 z + \cdots + g_n z^n$ we have

$$T_f T_g = T_g T_f, \qquad H_f Z H_g = H_g Z H_f, \tag{3}$$

where the last equality follows from $(ZH_f)(ZH_g) = (ZH_g)(ZH_f)$.
For $n \in \mathbb{N}$ and $z \in \mathbb{C}$ we denote by $V_n(z)$ the $n \times 1$ column vector

$$V_n(z) = \begin{bmatrix} 1 & z & \cdots & z^{n-1} \end{bmatrix}^\top .$$

This notation is convenient as it provides a compact way of writing polynomials. For example, a polynomial $a(z, w)$ in two variables z and w can be written as:

$$a(z, w) = \sum_{j,k=0}^{n-1} a_{jk} z^j w^k = V_n(z)^\top A V_n(w),$$

where A is the $n \times n$ coefficient matrix $\begin{bmatrix} a_{jk} \end{bmatrix}_{j,k=0}^{n-1}$ of $a(z, w)$.
The *resultant* R of the polynomials f and g is the $2n \times 2n$ matrix given as a 2×2 block matrix:

$$R = \begin{bmatrix} T_f & ZH_f \\ T_g & ZH_g \end{bmatrix}. \tag{4}$$

Notice that the action of R on $V_{2n}(z)$ is particularly simple:

$$RV_{2n}(z) = \begin{bmatrix} f(z)V_n(z) & g(z)V_n(z) \end{bmatrix}^\top . \tag{5}$$

Next we define the Bezoutian B of f and g. First consider the polynomial $f(z)g(w) - f(w)g(z)$ in two variables. Since this polynomial vanishes for all $w = z \in \mathbb{C}$, there exists a polynomial $b(z, w)$ in two variables such that

$$f(z)g(w) - f(w)g(z) = (z - w)b(z, w) \quad \text{for all} \quad z, w \in \mathbb{C}.$$

The *Bezoutian* B of f and g is the $n \times n$ coefficient matrix of $b(z, w)$:

$$b(z, w) = V_n(z)^\top B V_n(w), \quad z, w \in \mathbb{C}. \tag{6}$$

The null space or kernel of a matrix (or a linear transformation) A is denoted by $\ker A$. Its dimension is called the *nullity* of A.

2 A connection between R and B

To establish a connection between R and B we consider the polynomial $(z^n - w^n)b(z, w)$ and we find two ways of representing its coefficient matrix. To find the first representation we use the standard identity

$$z^n - w^n = (z - w) \sum_{j=0}^{n-1} z^{n-1-j} w^j = (z - w)V_n(z)^\top Z V_n(w), \quad z, w \in \mathbb{C},$$

matrix algebra and (5):

$$
\begin{aligned}
(z^n - w^n)b(z,w) &= (z-w)b(z,w)V_n(z)^\top Z V_n(w) \\
&= \big(f(z)g(w) - g(z)f(w)\big)V_n(z)^\top Z V_n(w) \\
&= \begin{bmatrix} f(z)V_n(z) \\ g(z)V_n(z) \end{bmatrix}^\top \begin{bmatrix} g(w)ZV_n(w) \\ -f(w)ZV_n(w) \end{bmatrix} \\
&= \begin{bmatrix} f(z)V_n(z) \\ g(z)V_n(z) \end{bmatrix}^\top \begin{bmatrix} 0 & Z \\ -Z & 0 \end{bmatrix} \begin{bmatrix} f(w)V_n(w) \\ g(w)V_n(w) \end{bmatrix} \\
&= V_{2n}(z)^\top R^\top \begin{bmatrix} 0 & Z \\ -Z & 0 \end{bmatrix} R\, V_{2n}(w).
\end{aligned}
$$

The second representation involves the Bezoutian:

$$
\begin{aligned}
(z^n - w^n)b(z,w) &= (z^n - w^n)V_n(z)^\top B\, V_n(w) \\
&= \big(z^n V_n(z)\big)^\top B\, V_n(w) - V_n(z)^\top B\,\big(w^n V_n(w)\big) \\
&= V_{2n}(z)^\top \begin{bmatrix} 0 & 0 \\ B & 0 \end{bmatrix} V_{2n}(w) + V_{2n}(z)^\top \begin{bmatrix} 0 & -B \\ 0 & 0 \end{bmatrix} V_{2n}(w) \\
&= V_{2n}(z)^\top \begin{bmatrix} 0 & -B \\ B & 0 \end{bmatrix} V_{2n}(w).
\end{aligned}
$$

These two representations of the coefficient matrix of $(z^n - w^n)b(z,w)$ provide a connection between R and B:

$$
R^\top \begin{bmatrix} 0 & Z \\ -Z & 0 \end{bmatrix} R = \begin{bmatrix} 0 & -B \\ B & 0 \end{bmatrix}. \tag{7}
$$

On the other hand, using the definition of R, (2) and (3) we obtain

$$
\begin{aligned}
R^\top \begin{bmatrix} 0 & Z \\ -Z & 0 \end{bmatrix} R &= \begin{bmatrix} ZT_f Z & ZT_g Z \\ H_f Z & H_g Z \end{bmatrix} \begin{bmatrix} ZT_g & H_g \\ -ZT_f & -H_f \end{bmatrix} \\
&= \begin{bmatrix} ZT_fT_g - ZT_gT_f & ZT_fZH_g - ZT_gZH_f \\ H_fT_g - H_gT_f & H_fZH_g - H_gZH_f \end{bmatrix} \\
&= \begin{bmatrix} 0 & -\big(H_fT_g - H_gT_f\big)^\top \\ H_fT_g - H_gT_f & 0 \end{bmatrix}.
\end{aligned}
$$

Together with (7), the last equality yields

$$
\begin{bmatrix} 0 & -\big(H_fT_g - H_gT_f\big)^\top \\ H_fT_g - H_gT_f & 0 \end{bmatrix} = \begin{bmatrix} 0 & -B \\ B & 0 \end{bmatrix},
$$

and thus

$$
B = H_fT_g - H_gT_f = B^\top. \tag{8}
$$

3 R and B have the same nullity

Equation (7) indicates that there is a connection between $\ker R$ and $\ker B$. An even more direct connection between $\ker R$ and $\ker B$ is obtained from (4), (8) and (3) (listed in the order in which they are used) as follows:

$$\begin{aligned}
\begin{bmatrix} I & 0 \\ T_f & ZH_f \end{bmatrix} R &= \begin{bmatrix} T_f & ZH_f \\ T_f^2 + ZH_f T_g & T_f ZH_f + ZH_f ZH_g \end{bmatrix} \\
&= \begin{bmatrix} T_f & ZH_f \\ ZB + \big(T_f + ZH_g\big)T_f & \big(T_f + ZH_g\big)ZH_f \end{bmatrix} \\
&= \begin{bmatrix} 0 & I \\ ZB & T_f + ZH_g \end{bmatrix} \begin{bmatrix} I & 0 \\ T_f & ZH_f \end{bmatrix} \\
&= \begin{bmatrix} 0 & I \\ Z & T_f + ZH_g \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ T_f & ZH_f \end{bmatrix}.
\end{aligned} \tag{9}$$

If we assume that $n = \deg f$, then H_f is invertible, yielding that the first (as well as the last) block matrix in (9) is invertible. Since the block matrix in (9) whose antidiagonal entries are I and Z is also invertible, (9) implies that R and B have the same nullities:

$$\dim \ker R = \dim \ker B. \tag{10}$$

4 The nullity of B in terms of f and g

Consider the multiplication operator

$$M : \mathbb{C}^2[z]_{<n} \to \mathbb{C}[z]_{<2n}$$

defined by

$$M \begin{bmatrix} u \\ v \end{bmatrix} = fu + gv, \quad u, v \in \mathbb{C}[z]_{<n}.$$

For a characterization of the null space $\ker M$ of M in terms of f and g we need the greatest common divisor h of f and g, its degree $k = \deg h$ and factorizations $f = \hat{f}h$, $g = \hat{g}h$. Then

$$\ker M = \left\{ \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{C}^2[z] \,:\, u = -\hat{g}q,\ v = \hat{f}q,\ q \in \mathbb{C}[z]_{<k} \right\}. \tag{11}$$

The inclusion $\supseteq$ in (11) is clear. To prove $\subseteq$, let $u, v \in \mathbb{C}[z]_{<n}$ and $\begin{bmatrix} u & v \end{bmatrix}^\top \in \ker M$. Then $fu + gv = 0$, implies $\hat{f}u = -\hat{g}v$. Since $\hat{f}$ and $\hat{g}$ have no common zeros, the last identity yields that there exist polynomials p and q such that $u = \hat{g}p$ and $v = \hat{f}q$. Substituting back to $\hat{f}u = -\hat{g}v$, we get $\hat{f}\hat{g}p = -\hat{g}\hat{f}q$. Hence $p = -q$. Since $\deg v < n$ and $\deg \hat{f} = n - k$, $v = \hat{f}q$ implies $\deg q < k$. This proves (11).

The standard basis for $\mathbb{C}^2[z]_{<n}$ is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} z^{n-1} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ z^{n-1} \end{bmatrix},$$

while the standard basis for $\mathbb{C}[z]_{<2n}$ is

$$1, z, \ldots, z^{n-1}, z^n, \ldots, z^{2n-1}.$$

The matrix representation for M with respect to these standard bases is $R^\top$, see (5). Therefore the nullity of $R^\top$ is $\dim\ker M$. Since (11) yields $\dim\ker M = k$ and the nullity of $R^\top$ equals the nullity of R, we have proved that the nullity of R is k and, by (10),

$$\dim\ker B = \dim\ker R = k = \deg h.$$

5 Final remarks

It was remarked in [6, p. 318] that the Bezoutian of a pair of polynomials is defined whenever $n \geq \max\{\deg f, \deg g\}$. We add to this that the same is true for the resultant and that if $n \geq m := \max\{\deg f, \deg g\}$, then formula (1) has to be replaced by the formula

$$\dim\ker B_n = \dim\ker R_{2n} = n - m + \deg\gcd(f, g), \tag{12}$$

where, for example, the index n in B_n indicates that B_n has size $n \times n$. Indeed, (12) follows from (1) and from the equalities

$$\dim\ker B_n = n - m + \dim\ker B_m \quad \text{and} \quad \dim\ker R_{2n} = n - m + \dim\ker R_{2m}.$$

The first of these two equalities holds because of (6), which implies $B_n = \begin{bmatrix} B_m & 0 \\ 0 & 0 \end{bmatrix}$, and the second follows from the reasoning in Section 4 with k in (11) replaced by $n - m + k$. Finally we note that (12) can be expressed as

$$\dim\ker B_n = \dim\ker R_n = \deg\gcd(\bar{f}, \bar{g}),$$

where

$$\bar{f}(y, z) = f_0 y^n + f_1 y^{n-1} z + \cdots + f_n z^n \quad \text{and} \quad \bar{g}(y, z) = g_0 y^n + g_1 y^{n-1} z + \cdots + g_n z^n$$

are homogenizations (in the sense of [7, pages 6–7]) of f and g, respectively. If $n > m = \max\{\deg f, \deg g\}$, then $y = 0$ is a common zero of $\bar{f}$ and $\bar{g}$ of multiplicity $n - m$. Since the zero $y = 0$ of the homogenization is commonly viewed as a "zero at infinity" of the original polynomial (see for example [2, 4.4.3]) we can now formulate the main theorem on the "generalized" Bezoutian B_n: The nullity of the Bezoutian matrix B_n associated with a pair of polynomials f and g equals the number of their common zeros including the "zero at infinity" and counting multiplicities.

References

[1] D.S. Bernstein, Matrix Mathematics: Theory, Facts, and Formulas. Second Edition, Princeton University Press, 2009.

[2] G. Blekherman, Nonnegative polynomials and sums of squares, in G. Blekherman, P.A. Parrilo, R.R. Thomas, eds., Semidefinite Optimization and Convex Algebraic Geometry, MOS-SIAM Series on Optimization, 13. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013, 159–202.

[3] H. Dym, Linear algebra in action. Graduate Studies in Mathematics, 78. American Mathematical Society, Providence, RI, 2007.

[4] M. Fiedler, Special Matrices and Their Applications in Numerical Mathematics. Second Edition, Dover Books on Mathematics, Mineola, NY, 2008.

[5] P.A. Fuhrmann, A polynomial approach to linear algebra. Second edition. Universitext. Springer, New York, 2012.

[6] D. Hinrichsen, A.J. Pritchard, Mathematical systems theory I. Modelling, state space analysis, stability and robustness. Texts in Applied Mathematics, 48. Springer, Heidelberg, 2010.

[7] M. Marshall, Positive polynomials and sums of squares. Mathematical Surveys and Monographs, 146. American Mathematical Society, Providence, RI, 2008.

[8] H.K. Wimmer, On the history of the Bezoutian and the resultant matrix. Linear Algebra Appl. 128 (1990), 27–34.

Branko Ćurgus
Department of Mathematics
Western Washington University
Bellingham, WA 98225, USA
e-mail: curgus@wwu.edu

Aad Dijksma
Johann Bernoulli Institute
of Mathematics and Computer Science
University of Groningen
P.O. Box 407
9700 AK Groningen, The Netherlands
e-mail: a.dijksma@rug.nl