Matrix Vieta Theorem

Dmitry FUCHS and Albert SCHWARZ*
Department of Mathematics
University of California
Davis Ca 95616, USA

1. Introduction

Consider a matrix algebraic equation

$$X^n + A_1 X^{n-1} + \ldots + A_n = 0, \tag{1}$$

where the coefficients $A_1, \ldots, A_n$ as well as solutions X are supposed to be square complex matrices of some order k. For an usual algebraic equation of degree n the classical Vieta formulas express the coefficients in terms of the n solutions. However, a matrix degree n equation generically has $\binom{nk}{k}$ rather than n solutions. (Throughout this article the word *generic* refers to a Zariski open set.)

We call n solutions $X_1, \ldots, X_n$ of the equation (1) *independent* if they determine the coefficients $A_1, \ldots, A_n$ (a more technical explanation of independence see in Section 2). However, the expressions of $A_1, \ldots, A_n$ in terms of $X_1, \ldots, X_n$ are much less elegant, than Vieta formulas (see, e.g. formulas (3) – (5) below). Still there are some relations between A's and X's which are very similar to Vieta.

Theorem 1.1. *If solutions $X_1, \ldots, X_n$ of the equation* (1) *are independent, then*

$$\begin{aligned} \operatorname{tr} A_1 &= -(\operatorname{tr} X_1 + \ldots + \operatorname{tr} X_n), \\ \det A_n &= (-1)^{nk} \det X_1 \ldots \det X_n. \end{aligned} \tag{2}$$

Theorem 1.1 is proved in Section 4 (a more direct proof for $n = 2$ is given in Section 3). In Section 5 we discuss a generalization of Theorem 1.1 from complex matrix algebras to arbitrary associative unitary rings.

Theorem 1.1 is elementary, however it is connected with some constructions of modern Mathematics. For every associative algebra A one can construct a linear space $F(A) = A^2/[A, A]$. This space appeared in [K] as the space of 0- forms on a non-commutative formal manifold. (Such a manifold is determined by a free associative algebra A. For the free associative algebra A with generators $a_1, \ldots, a_n$ the space $F(A)$ is spanned by cyclic words of length ≥ 2 with letters $a_1, \ldots a_n$.) The space $F(A)$ appeared also in [GS] and [AW] in relation to other problems. It is easy to understand that the formula for $\operatorname{tr} A_1$ can be interpreted as a non-trivial identity in $F(A)$.

* This work was partially supported by NSF grant DMS-9201366. Research at MSRI was supported by NSF grant DMS-9022140.

2. Independent matrices.

DEFINITION 2.1. Matrices $X_1, \dots, X_n$ are called *independent* if the bloc Vandermonde determinant is not zero:

$$\begin{vmatrix} I & I & \dots & I \\ X_1 & X_2 & \dots & X_n \\ \dots & \dots & \dots & \dots \\ X_1^{n-1} & X_2^{n-1} & \dots & X_n^{n-1} \end{vmatrix} \neq 0.$$

For $n = 2$ the independence condition means that $\det(X_1 - X_2) \neq 0$. For $n \geq 3$ it does not imply and is not implied by the condition $\det(X_i - X_j) \neq 0$, $1 \leq i < j \leq n$.

It is obvious that a generic equation (1) has n independent solutions (otherwise the above determinant would have vanished for any n solutions of any equation (1)). It is clear also that the matrices $X_1, \dots, X_n$ are independent if and only is there exist unique $A_1, \dots, A_n$ such that $X_1, \dots, X_n$ satisfy the equation (1). In other words, for independent $X_1, \dots, X_n$ the matrices $A_1, \dots, A_n$ may be expressed via $X_1, \dots, X_n$. For example, if $n = 2$ then

$$\begin{aligned} A_1 &= -(X_1^2 - X_2^2)(X_1 - X_2)^{-1}, \\ A_2 &= -X_1^2 + (X_1^2 - X_2^2)(X_1 - X_2)^{-1} X_1. \end{aligned} \tag{3}$$

For $n \geq 3$ it is impossible to write a formula valid for all independent matrices; for example, if $n = 3$, then

$$\begin{aligned} A_1 = -((X_1^3 - X_2^3)(X_1 - X_2)^{-1} - (X_1^3 - X_3^3)(X_1 - X_3)^{-1}) \\ ((X_1^2 - X_2^2)(X_1 - X_2)^{-1} - (X_1^2 - X_3^2)(X_1 - X_3)^{-1})^{-1} \end{aligned} \tag{4}$$

provided that the right hand side exists (which does not follow from X_1, X_2, X_3 being independent); otherwise the expression will be different.

An expression of $A_1, \dots, A_n$ as functions of $X_1, \dots, X_n$, valid for generic independent $X_1, \dots, X_n$, may be given in terms of Gelfand–Retakh's quasideterminants, which are defined as follows.

DEFINITION 2.2 [GR]. Let $A = \{\|a_{ij}\|, i \in I, j \in J\}$ be a square matrix of order $n = \operatorname{card} I = \operatorname{card} J$ with formal non-commutative entries a_{ij}. For $p \in I, q \in J$ denote by A^{pq} the submatrix $\{\|a_{ij}\|, i \in I - p, j \in J - q\}$ of A. The formula

$$|A|_{pq} = a_{pq} - \sum_{\substack{i \in I - p \\ j \in J - q}} a_{pj} |A^{pq}|_{ij}^{-1} a_{ip}$$

(which reduces to $|A|_{pq} = a_{pq}$ if $n = 1$) defines inductively n^2 *quasideterminants* $|A|_{pq}$ of the matrix A. (In the commutative case $|A|_{pq} = \pm \det A / \det A_{pq}$.)

Quasideterminants possess some basic properties of determinants; in particular, the following *Kramer rule* holds.

PROPOSITION 2.3 [GR]. *If $(x_1, \dots, x_n)$ is the solution of a system of equations*

$$\sum_{j=1}^{n} a_{ij} x_j = \xi_i \; (i = 1, \dots, n),$$

then for any i

$$x_j = |A|_{ij}^{-1}|A_j(\xi)|_{ij},$$

where $A = \|a_{ij}\|$ and $A_j(\xi)$ is obtained from A be replacing its j-th column by the column $(\xi_1, \dots, \xi_n)$.

COROLLARY 2.4. *For generic independent $X_1, \dots, X_n$ and for arbitrary i*

$$-A_j =$$

$$\begin{vmatrix} I & X_1 & \dots & X_1^{n-1} \\ I & X_2 & \dots & X_2^{n-1} \\ \dots & \dots & \dots & \dots \\ I & X_n & \dots & X_n^{n-1} \end{vmatrix}_{i,n-j}^{-1} \begin{vmatrix} I & X_1 & \dots & X_1^{n-j-1} & X_1^n & X_1^{n-j+1} & \dots & X_1^{n-1} \\ I & X_2 & \dots & X_2^{n-j-1} & X_2^n & X_2^{n-j+1} & \dots & X_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ I & X_n & \dots & X_n^{n-j-1} & X_n^n & X_n^{n-j+1} & \dots & X_n^{n-1} \end{vmatrix}_{i,n-j} \quad (5)$$

This expression times $(-1)^j$ (with a minor change of notations and with $i = n$) is called in [GKLLRT], Section 7.1, the j-th elementary symmetric function in $X_1, \dots, X_n$. It is proved in [GKLLRT] and is obvious from Corollary 2.4, that for generic $X_1, \dots, X_n$ it is really symmetric in $X_1, \dots, X_n$.

3. The case $n = 2$

Since for any square matrices X, Y

$$\begin{vmatrix} I & I \\ X & Y \end{vmatrix} = \det(Y - X),$$

the following is precisely Theorem 1.1 for the case $n = 2$.

PROPOSITION 3.1. *Let A, B, X, Y be square matrices of the same order with $X - Y$ being non-degenerate. If*

$$\begin{aligned} X^2 + AX + B &= 0, \\ Y^2 + AY + B &= 0. \end{aligned} \quad (6)$$

Then

$$\begin{aligned} \operatorname{tr} A &= -(\operatorname{tr} X + \operatorname{tr} Y) \\ \det B &= \det X \det Y. \end{aligned}$$

PROOF. 1. The equalities (6) imply

$$-A = (X^2 - Y^2)(X - Y)^{-1}, \quad (7)$$

and since $X^2 - Y^2 = (X + Y)(X - Y) + XY - YX$, then

$$-A = X + Y + (XY - YX)(X - Y)^{-1}.$$

Therefore

$$-\operatorname{tr} A = \operatorname{tr} X + \operatorname{tr} Y + \operatorname{tr} XY(X - Y)^{-1} - \operatorname{tr} YX(X - Y)^{-1}, \quad (8)$$

and the relations
$$X(X-Y)^{-1} = Y(X-Y)^{-1} + I,$$
$$(X-Y)^{-1}X = (X-Y)^{-1}Y + I$$
together with the identity $\mathrm{tr}\, UV = \mathrm{tr}\, VU$ imply

$$\mathrm{tr}\, XY(X-Y)^{-1} = \mathrm{tr}\, Y(X-Y)^{-1}X = \mathrm{tr}\, Y(X-Y)^{-1}Y + \mathrm{tr}\, Y,$$
$$\mathrm{tr}\, YX(X-Y)^{-1} = \mathrm{tr}\, X(X-Y)^{-1}Y = \mathrm{tr}\, Y(X-Y)^{-1}Y + \mathrm{tr}\, Y.$$

Hence $\mathrm{tr}\, XY(X-Y)^{-1} = \mathrm{tr}\, YX(X-Y)^{-1}$, and (8) yields

$$-\mathrm{tr}\, A = \mathrm{tr}\, X + \mathrm{tr}\, Y.$$

2. The first of the equalities (6) implies

$$B = -(X+A)X, \tag{9}$$

and using (7) we get

$$\begin{aligned}(X+A)(X-Y) &= (X-(X^2-Y^2)(X-Y)^{-1})(X-Y)\\ &= X^2 - XY - X^2 + Y^2\\ &= Y^2 - XY = (X-Y)(-Y).\end{aligned}$$

Therefore
$$X + A = (X-Y)(-Y)(X-Y)^{-1},$$
$$\det(X+A) = \det(-Y),$$
and in virtue of (9)

$$\det B = \det(-(X+A))\det X = \det Y \det X.$$

4. The general case

Unlike the above proof for $n=2$, our proof in the general case is not reduced to a direct calculation and uses the specifics of the matrix algebra.

LEMMA 4.1. *Let $A_1, \ldots, A_n, X_1, \ldots, X_n$ be square matrices of some order k, and let the eigenvalues of the matrices $X_1, \ldots, X_n$ be kn pairwise different complex numbers. If*

$$X_i^n + A_1 X_i^{n-1} + \ldots + A_n = 0 \tag{10}$$

for $i = 1, \ldots, n$, then
$$\mathrm{tr}\, A_1 = -(\mathrm{tr}\, X_1 + \ldots + \mathrm{tr}\, X_n),$$
$$\det A_n = (-1)^{kn} \det X_1 \ldots \det X_n.$$

Proof. Let λ_{ij}, $j = 1, \ldots, k$ be eigenvalues of X_i and let $X_i v_{ij} = \lambda_{ij} v_{ij}$, $v_{ij} \neq 0$. Then for arbitrary i, j (10) implies

$$\begin{aligned} 0 &= (X_i^n + A_1 X_i^{n-1} + \ldots + A_n) v_{ij} \\ &= (\lambda_{ij}^n + \lambda_{ij}^{n-1} A_1 + \ldots + A_n) v_{ij}, \end{aligned}$$

whence

$$\det(\lambda_{ij}^n + \lambda_{ij}^{n-1} A_1 + \ldots + A_n) = 0.$$

Obviously,

$$P(\lambda) = \det(\lambda^n + \lambda^{n-1} A_1 + \ldots + A_n)$$

is a monic polynomial in λ of degree kn, and since $P(\lambda_{ij}) = 0$ and all λ_{ij} are different, then

$$\det(\lambda^n + \lambda^{n-1} A_1 + \ldots + A_n) = \prod_{i,j} (\lambda - \lambda_{ij}). \tag{11}$$

Equate constant terms and the coefficients in the term with λ^{kn-1} of the two sides of (11); we have

$$\begin{aligned} \det A_n &= \prod_{i,j} (-\lambda_{ij}) = (-1)^{kn} \prod_{i,j} \lambda_{ij} = (-1)^{kn} \prod_i \det X_i. \\ \mathrm{tr} A_1 &= \sum_{i,j} (-\lambda_{ij}) = -\sum_i \mathrm{tr}\, A_i. \end{aligned}$$

Remark 4.2. Equating other coefficients of polynomials in (11), we may get $kn - 2$ more identities. Most of them involve matrices, obtained by combination of columns of different matrices A_i, but in the two extreme cases (those of λ^p with $p = kn - 2$ and 1) we get the formulas which are worth mentioning:

$$\mathrm{tr}\, A_2 + \sigma_2(A_1) = \sum_{1 \leq i < j \leq n} (\mathrm{tr}\, X_i \mathrm{tr}\, X_j) + \sum_{1 \leq i \leq n} \sigma_2(X_i); \tag{12}$$

$$\det A_{n-1} \cdot \mathrm{tr}\, A_n A_{n-1}^{-1} = (-1)^{kn-1} (\prod \det X_i)(\sum \mathrm{tr}\, X_i^{-1}), \tag{13}$$

where σ_2 in (12) denotes the second coefficient (the coefficient in the term with λ^{k-2}) of the characteristic polynomial of a matrix, and (13) is valid only if $A_{n-1}, X_1, \ldots, X_n$ are non-degenerate.

Proof of Theorem 1.1. Let $U, V \subset \mathrm{Mat}_k\mathbb{C} \times \ldots \times \mathrm{Mat}_k\mathbb{C}$ (n factors) are respectively the set of all n-tuples $(X_1, \ldots, X_n)$ of independent matrices and the set of all n-tuples $(X_1, \ldots, X_n)$ of matrices whose eigenvalues are kn pairwise different complex numbers. Obviously the both sets are Zariski open and non-empty, hence $U \cap V$ is dense in U. According to Lemma, the equalities (2) both hold in $U \cap V$. Since the both sides of each of the equalities (2) are continuous on U (with respect to $X_1, \ldots, X_n$), the equalities (2) hold on the whole U.

5. A further generalization

Let R be an associative ring with unity, and k be a field. Suppose that there fixed either an additive homomorphism $\mathrm{tr}\colon R \to k$ satisfying the condition $\mathrm{tr}UV = \mathrm{tr}VU$ for any $U, V \in R$, or a ring homomorphism $\det\colon R \to k$ (or both).

PROPOSITION 5.1. *Let $A, B, X, Y \in R$ with $X - Y$ being invertible. If $X^2 + AX + B = 0, Y^2 + AY + B = 0$, then $\mathrm{tr}\, A = -(\mathrm{tr}\, X + \mathrm{tr}\, Y)$ and/or $\det B = \det X \det Y$, whichever is defined.*

The proof is the same as in Section 3.

To generalize to arbitrary rings the general case of Theorem 1.1 we need an explicit expression of A_1, A_n via $X_1, \ldots, X_n$.

THEOREM 5.2. *Let*

$$A_1 = a_1(X_1, \ldots, X_n),$$
$$A_n = a_n(X_1, \ldots, X_n)$$

be expressions of A_1, A_n via $X_1, \ldots, X_n$ involving ring operations and taking inverses and valid where these inverses exist (like (3), (4), (5)). *Then for any $X_1, \ldots, X_n \in R$ each of the equalities*

$$\mathrm{tr}\, a_1(X_1, \ldots, X_n) = -(\mathrm{tr}\, X_1 + \ldots + \mathrm{tr}\, X_n),$$
$$\det a_n(X_1 \ldots, X_n) = \det(-X_1) \ldots \det(-X_n)$$

holds provided that the both sides exist (that is tr or det is defined and the inverses involved in a_1 or a_n exist in R).

Theorem 5.2 cannot be proved by arguments similar to that of Section 4. But it is known that identities which hold for complex matrices hold also in arbitrary associative rings with unities (see [A], Section 12.4.3). Hence Theorem 5.2 follows from Theorem 1.1.

Acknowledgement. We are indebted to C. Itzykson and I. Kaplansky for interesting discussion.

BIBLIOGRAPHY

[A] M. ARTIN. Algebra. Prentice Hall, Englewood Cliffs NJ, 1991.

[AW] S.L. ADLER, YONG-SHI WU. Algebraic and geometric aspects of generalized quantum dynamics. Preprint hep-th 9405054.

[GR] I.M. GELFAND, V.S. RETAKH. A theory of noncommutative determinants and characteristic functions of graphs, I. Publ. du LACIM, Univ. de Québec à Montréal, no. 14, P. 1–26.

[GKLLRT] I.M. GELFAND, D. KROB, A. LASCOUX, B. LECLERC, V.S. RETAKH, J.–Y. THIBON. Noncommutative symmetric functions. Preprint hep-th 9407124.

[GS] I.M. GELFAND, M.M. SMIRNOV. The algebra of Chern-Simons classes and the Poisson brackets on it. Preprint hep-th 9404103.

[K] M. KONTSEVICH. Formal (non)-commutative symplectic geometry. In: Gelfand Mathematical Seminar, 1992. Birkhäuser, Boston, 1993, P. 173–189.