Elem. Math. 75 (2020) 1 – 17
0013-6018/20/020001-17

Elemente der Mathematik

A matrix viewpoint for various algebraic extensions

Gene Abrams and P.N. Ánh

Gene Abrams is Professor of Mathematics at University of Colorado, Colorado Springs. He is the author of more than fifty research articles, as well as a coauthor of the text *Leavitt path algebras* (Springer LNM Volume 2191). He has also been recognized for various teaching and community outreach activities.

Except for a few years spent teaching in Vietnam, P. N. Ánh has done his research at Alfréd Rényi Institute of Mathematics in Hungary. He enjoys teaching at all levels, including both elementary school and high school. He believes strongly in the beauty and unity of universal mathematics; math should not be just an umbrella to cover many disparate ideas.

A standard, central result which is presented in a first course on rings is the following: Let K be a field and let $f(x)$ be a nonconstant polynomial in $K[x]$. Then there exists an extension field F of K and an element $\alpha \in F$ such that $f(\alpha) = 0$.[1] Most approaches to establishing this result follow this outline: it is sufficient to assume that $f(x)$ is irreducible in $K[x]$; form the quotient ring $F = K[x]/f(x)K[x]$, which is a field by the irreducibility of $f(x)$; view K as a subfield of F; show that $\alpha = x + f(x)K[x] \in F$ has $f(\alpha) = 0$. While of course this approach to constructing zeros of polynomials is both *mathematically*

[1] This result is called 'Kroneker's Theorem (Basic Goal)' in [2].

Matrizen sind einfache aber faszinierende mathematische Strukturen mit einer Vielzahl von Anwendungen und Interpretationen. Eine der wichtigsten Ideen besteht darin, zu einer Matrix verschiedene Polynome zu assoziieren (z.B. das Minimalpolynom oder das charakteristische Polynom). Im vorliegenden Artikel wird gezeigt, dass auch der umgekehrte Prozess einige äusserst nützliche Einsichten liefern kann. Man beginnt mit einem Polynom und ordnet ihm eine Matrix (die Begleitmatrix) zu. Dieser Matrixansatz bietet sodann eine transparentere, intuitivere Art und Weise, die Existenz und den Aufbau der Wurzeln des gegebenen Polynoms zu verstehen. Dieser Ansatz verbindet Matrizen auch mit mehreren interessanten zahlentheoretischen Problemen, wie z.B. mit den Mersenne-Primzahlen und mit der Primfaktorzerlegung grosser ganzer Zahlen. Am Ende zeigt es sich, dass Matrizen, das Lösen von Polynomgleichungen und endliche Körpererweiterungen verschiedene Gesichter desselben Juwels sind.

sound and *aesthetically pleasing*, many students find the quotient ring approach to be both *mathematically challenging* and *aesthetically unpleasant*.

Furthermore, the approach taken in most such introductory ring theory courses to establish the existence of an algebraic closure of a field K, via a Zorn Lemma argument on the set of all algebraic extension fields of K, strikes most of the authors' students as being, at best, *mysterious*. See, e.g., [2, Theorem 31.17].

On the other hand, given any nonzero monic polynomial $f(x) \in K[x]$ of degree $n \geq 1$, one may form the usual *companion matrix* A_f *of* $f(x)$, an $n \times n$ matrix over K (see Section 1). It is well-known that $f(A_f) = 0$ in the matrix ring $M_n(K)$ (for completeness we provide a proof below). With this as motivation, we assert that the matrix context provides an approach to finding zeros of polynomials in field extensions which students may find more appealing and intuitively clearer than the quotient ring approach. In a similar way, we show below in Theorem 3.1 that a matrix context can be used to more concretely construct an algebraic closure of a field.

The previous remarks point to just two places where a matrix approach can be used in the setting of algebraic extensions. Our goal in this note is to call attention to the ubiquity of places (over and above these two) where the more traditional approach to constructing field extensions may be replaced by a (perhaps more user friendly) matrix-centered approach. Such additional places include the construction of finite fields and the parameterization of algebraic curves, which we mention in the final section.

We note that viewing algebraic elements as linear transformations (i.e., as companion matrices) allows the successful and effective use of well-known notions of linear algebra (e.g., traces, determinants) in algebraic number theory; see, e.g., Eichler's classic text [1]. All this notwithstanding, it remains somewhat of a mystery to the authors as to why the matrix approach has not been more extensively used to develop the theory of algebraic field extensions.

A few words about terminology. All fields are commutative. Rings are associative with identity. A ring is a *division ring* or a *skew field* if all nonzero elements are (two-sided) units. For a ring R, $M_n(R)$ denotes the ring of $n \times n$ matrices with entries in R. Two elements $M, N \in M_n(R)$ are *similar* in case there exists an invertible $S \in M_n(R)$ with $M = S^{-1}NS$.

For a ring R, $R[x]$ denotes the ring of polynomials in the central variable x with coefficients in R. For $M \in M_n(R)$ and $f(x) = \sum_{i=0}^{n} a_i x^i \in R[x]$, the expression $f(M)$ denotes the element $\sum_{i=0}^{n} a_i M^i \in M_n(R)$. We emphasize that for the constant term $a_0 = a_0 x^0$ of $f(x)$, the matrix $a_0 M^0$ is interpreted as the diagonal (scalar) matrix with a_0 in each of the main diagonal entries, and 0 elsewhere; rephrased, $a_0 M^0$ is defined to be $a_0 I_n$.

Remark 0.1. Already we encounter a difference between the two approaches. Eventually we must interpret how the field K is to be embedded in a larger extension field in which there is a zero of $f(x)$. In the standard "quotient of $K[x]$" approach, K appears as cosets of the form $k + f(x)K[x]$ with $k \in K$. In the matrix approach, the field K appears as

scalar matrices in $M_n(K)$. This second context seems more intuitive for students on first encounter.

For the sake of self-containedness and elementary introduction we provide proofs of a few key well-known results, most of which are presented in a way to be understood by readers even with modest background knowledge.[2]

1 Constructing zeros of polynomials in extension fields

In this first section we achieve the article's first goal: starting with a non-constant monic polynomial $f(x) \in K[x]$, build a "concrete" extension field F of K which contains a zero of $f(x)$.

So we start with a monic polynomial $f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in R[x]$ $(n \geq 1)$. The matrix

$$A_f = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix} \in M_n(R)$$

is called the *companion matrix* of f. If $f(x)$ is a monic linear polynomial (i.e., of the form $x - r$ for $r \in R$), then the companion "matrix" is just r itself.

The following result is classical; it is a particular case of the famous Cayley–Hamilton Theorem in matrix theory.

Proposition 1.1. *Let R be a commutative ring. Let $f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in R[x]$ $(n \geq 1)$. Then A_f is a solution of f; that is, $f(A_f) = 0$ in $M_n(R)$.*

Proof. For $1 \leq i \leq n$ we let $\vec{e}_i$ denote the standard $n \times 1$ column vector in R^n with 1 in the ith row and zeros elsewhere. For simplicity we denote A_f by A. Then easy matrix multiplication gives $A\vec{e}_1 = \vec{e}_2, \ldots, A\vec{e}_{n-1} = \vec{e}_n$, $A\vec{e}_n = \sum_{i=0}^{n-1} -\vec{e}_{i+1} a_i$. In particular this gives $A^i \vec{e}_1 = \vec{e}_{i+1}$ (for $1 \leq i < n$), which then easily yields $(A^n + \sum_{i=0}^{n-1} a_i A^i)\vec{e}_1 = \vec{0}$ (the $n \times 1$ zero column vector). Consequently, for each $0 \leq j \leq n-1$ we have

$$\left(A^n + \sum_{i=0}^{n-1} a_i A^i\right)\vec{e}_{j+1} = \left(A^n + \sum_{i=0}^{n-1} a_i A^i\right)(A^j \vec{e}_1) = A^j \left[\left(A^n + \sum_{i=0}^{n-1} a_i A^i\right)\vec{e}_1\right] = \vec{0}$$

[2]The authors are grateful to Tamás Szamuely and to the referees for their comments and remarks improving the presentation.

(we note that the commutativity of R is used in the penultimate equality). So the matrix $f(A) = A^n + \sum_{i=0}^{n-1} a_i A^i$ sends every basis vector $\vec{e}_j$ $(1 \leq j \leq n)$ of R^n to zero, and so this matrix is the zero matrix, thus giving the result. □

The columns of A_f are easily seen to form a basis of R^n if and only if a_0 is invertible, so that the invertibility of A_f is exactly determined by the invertibility of a_0. More specifically, since $A_f^{-1}(\vec{e}_{i+1}) = A_f^{-1}A_f(\vec{e}_i) = \vec{e}_i$ for all $0 < i < n$, we have immediately that $A_f^{-1}(\vec{e}_i) = \vec{e}_{i-1}$ for all $i > 1$, and $\vec{e}_n = A_f^{-1}A_f(\vec{e}_n) = -A_f^{-1}(\sum_{i=1}^{n} \vec{e}_i a_{i-1}) = -(A_f^{-1}\vec{e}_1 a_0 + \sum_{i=1}^{n-1} \vec{e}_i a_i)$. So we get an explicit description of A_f^{-1}:

$$A_f^{-1} = \begin{pmatrix} -a_1a_0^{-1} & 1 & 0 & \cdots & 0 & 0 \\ -a_2a_0^{-1} & 0 & 1 & \cdots & 0 & 0 \\ -a_3a_0^{-1} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ -a_na_0^{-1} & 0 & \cdots & 0 & 0 & 1 \\ -a_0^{-1} & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} \in M_n(R).$$

Remark 1.2. Examples show that $f(A_f)$ need not be 0 in $M_n(R)$ when R is not commutative, even if R is a division ring. For instance, over the division ring $\mathbb{H}$ of quaternions, for $f(x) = x^2 - ix - j$ we have $A_f = \begin{pmatrix} 0 & j \\ 1 & i \end{pmatrix}$, and $f(A_f) = \begin{pmatrix} 0 & -2k \\ 0 & 0 \end{pmatrix} \neq 0$.

Example 1.3. Of course there are two classic examples.

If $K = \mathbb{Q}$ and $f(x) = x^2 - 2$, then $A_f = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$, and $A_f^2 - 2I_2 = 0$ in $M_2(\mathbb{Q})$; that is, A_f is a square root of 2.

If $K = \mathbb{R}$ and $f(x) = x^2 + 1$, then $A_f = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and $A_f^2 + 1I_2 = 0$ in $M_2(\mathbb{R})$; that is, A_f is a square root of -1.

Students may find the matrix representation of a zero of the two indicated polynomials to be more 'concrete' than the symbol $\sqrt{2}$ and (especially) the symbol i, respectively.

Again let $f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in K[x]$ (and assume $n \geq 2$). If $B \in M_n(K)$ has $f(B) = 0$ in $M_n(K)$ (i.e., if B is a root of $f(x)$ in $M_n(K)$) then we set

$$K[B] := \left\{ \sum_{i=0}^{n-1} k_i B^i \;\middle|\; k_i \in K \right\}.$$

Because $f(B) = 0$ immediately gives $B^n = -\sum_{i=0}^{n-1} a_i B^i$, $K[B]$ is easily seen to be closed under multiplication, and thus becomes a ring.

We now achieve the first goal of the article.

Theorem 1.4. *Let K be a field, and $f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in K[x]$ a monic irreducible polynomial. Let $B \in M_n(K)$ for which $f(B) = 0$ in $M_n(K)$; for instance, let $B = A_f$. Then $K[B]$ is a field extension of K.*

In particular, $K[A_f]$ is a field extension of K which contains a zero of $f(x)$.

Proof. Every non-zero element of $K[B]$ can be written (uniquely) as $g(B)$ for some $0 \neq g(x) \in K[x]$ having $\deg(g) < n$. Therefore $g(x)$ and $f(x)$ are coprime because $f(x)$ is irreducible and $\deg(g(x)) < \deg(f(x))$. Consequently, there exist $q(x), r(x) \in K[x]$ with $\deg(r(x)) < n$ satisfying $1 = g(x)r(x) + f(x)q(x)$; whence $1 = g(B)r(B) + f(B)q(B) = g(B)r(B)$. Thus every nonzero element of $K[B]$ is invertible, and so $K[B]$ is a field. □

Remark 1.5. We note that $K[B] \subsetneqq M_n(K)$ by a dimension argument ($n < n^2$). Since K is the center of $M_n(K)$ and $K \subsetneqq K[B]$, we see that there must be matrices $X \in M_n(K)$ which do not commute with B. Therefore, the subring of $M_n(K)$ generated by $F = K[B]$ and such an X provides handy examples of an extension of a ring F by a (subring $\mathbb{Z}[X]$ generated by a) single element X which are more intricate than the examples which arise as images of the usual polynomial rings. Such an approach to, and examples of, these types of ring extensions could make the concept of both polynomial rings (even skew polynomial rings) and free products of rings more accessible to beginners.

Although the observations that we will make in the remainder of this section are somewhat tangential to the goal achieved in Theorem 1.4, we believe it is both interesting and useful to obtain some additional properties of fields of the form $K[B]$ (where B is a zero in $M_n(K)$ of the monic irreducible polynomial $f(x) \in K[x]$).

The following idea is fairly standard. For $\Sigma \subseteq M_n(K)$, the *centralizer* of Σ is the set $Z(\Sigma) = \{N \in M_n(K) \mid NS = SN \text{ for all } S \in \Sigma\}$; i.e., the set of matrices which commute with every one of the elements of Σ. Because matrix multiplication is of course not commutative in general, for a given set $\Sigma \subseteq M_n(K)$ it is certainly possible to have $Z(\Sigma)$ properly contained in $M_n(K)$. For an easy example, if $\Sigma = M_n(K)$ then $Z(\Sigma) = K$.

For $\Sigma \subseteq M_n(K)$, we say that Σ has the *double centralizer property* in case $\Sigma = Z(Z(\Sigma))$; that is, in case $\Sigma = \{M \in M_n(K) \mid MN = NM \text{ for all } N \in Z(\Sigma)\}$. Note that by definition double centralizers are subalgebras, and all maximal commutative subrings of $M_n(K)$ have the double centralizer property; additionally, forming double centralizers is a closure operation. Originally, von Neumann invented the double centralizer property to define rings of operators which are called nowadays *von Neumann algebras*. We comment in a footnote below on the existence of subalgebras of $M_n(K)$ which do not have the double centralizer property.

Proposition 1.6. *Let K be a field, and $f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in K[x]$ a monic irreducible polynomial. Let $B, B' \in M_n(K)$ for which $f(B) = 0 = f(B')$ in $M_n(K)$. So, by Theorem* 1.4*, $K[B]$ (resp., $K[B']$) is a field.*

1. *$K[B] \cong K[B']$ as K-algebras; and each is isomorphic to $K[x]/f(x)K[x]$.*
2. *B is similar over K to B' (as each is similar in particular to A_f).*
3. *$K[B]$ is a maximal commutative subalgebra in $M_n(K)$, in particular a maximal subfield; that is, if E is a field with $K[B] \leq E \leq M_n(K)$, then $E = K[B]$.*
4. *$K[B]$ has the double centralizer property.*

Proof. (1) The map $\varphi : K[x] \to K[B]$, $x \mapsto B$, is easily shown to be a K-algebra surjection with kernel $f(x)K[x]$. The two necessary properties of B used to establish this isomorphism are that $f(x)$ is irreducible and that $f(B) = 0$; thus the identical proof gives $K[B'] \cong K[x]/f(x)K[x]$ as well.

(2) In particular, $E := K[A_f]$ is also a field, isomorphic to $F = K[B]$ via the K-isomorphism ϕ sending A_f to B. So K^n is a vector space over F; comparing the K-dimension of K^n to that of F (both are equal to n) implies that K^n is one dimensional over F. K^n admits another F- vector space structure via ϕ, by putting $g(B) \star v = \phi^{-1}(g(B))v$ for $g(x) \in K[x]$. Therefore there is a K-linear isomorphism $S : K^n \to K^n$, i.e., an invertible matrix $S \in M_n(K)$, such that $(BS)v = B(Sv) = S(B \star v) = S(A_f v) = SA_f v$ for every $v \in K^n$. Thus $SA_f = BS$, and so $A_f = S^{-1}BS$.

(3) If a matrix $X \in M_n(K)$ satisfies $XB = BX$, then X is an F-linear transformation of the one-dimensional F-vector space K^n, whence $X \in F = K[B]$ holds, i.e., F is a maximal commutative subring of $M_n(K)$.

(4) is an immediate consequence of (3). □

Example 1.7. When $f(x) = x^2 - 2 \in \mathbb{Q}[x]$, then $A_f = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$. Because in any field there are at most n zeros for a polynomial of degree n, we see that the (only) other zero of $f(x)$ in the field $\mathbb{Q}[A_f]$ is $-A_f = \begin{pmatrix} 0 & -2 \\ -1 & 0 \end{pmatrix}$. By Proposition 1.6 we have that $-A_f$ is similar to A_f; we note that the similarity matrix S is $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Remark 1.8. In contrast to the previously remarked behavior in a field, a polynomial of degree n may have more than n zeros in $M_n(K)$. For example, each of

$$A_f = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}, \quad -A_f = \begin{pmatrix} 0 & -2 \\ -1 & 0 \end{pmatrix}, \quad A_f^t = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}, \quad \text{and} \quad -A_f^t = \begin{pmatrix} 0 & -1 \\ -2 & 0 \end{pmatrix}$$

is a zero of $f(x) = x^2 - 2$ in $M_2(\mathbb{Q})$.

Proposition 1.6(2) says that any two zeros of a monic irreducible polynomial are similar.[3] This prompts the obvious followup question of whether the same statement holds for all

[3]The similarity of roots established in Proposition 1.6(2) is a special case of the classical Noether–Skolem Theorem on isomorphisms of simple subalgebras in matrix rings. The double centralizer property of Proposition 1.6(4) is a specific consequence of Riesz' theorem, which says that if A is an $n \times n$ matrix over a field K which commutes with every matrix which commutes with an $n \times n$ matrix B (i.e., if A is in the double centralizer of B), then A can be written as a polynomial in B with coefficients in K, whence $K[B]$ has the double centralizer property. Examples of subalgebras of $M_n(K)$ that do not have the double centralizer property are rings T_n

(not necessarily irreducible) polynomials. A moment's thought shows that this is clearly not true; for instance, for $f(x) = (x-1)(x-2) \in \mathbb{Q}[x]$, the zeros 1 and 2 are of course not conjugate over $M_2(\mathbb{Q})$. It is just as easy to produce counterexamples in $M_n(R)$ for any n.

2 A natural followup discussion: matrices similar to their transposes

Of course, for any (not necessarily irreducible) $f(x) \in K[x]$, if A is any matrix in $M_n(K)$ which is a root of $f(x)$, then the transpose A^t of A must also be a root of $f(x)$ as well (simply take $(\ )^t$ of both sides of the equation $f(A) = 0$). So in particular by Proposition 1.6(2), we conclude the following.

If $f(x)$ is monic irreducible in $K[x]$,
then the companion matrix A_f is similar to its transpose A_f^t.

Although Proposition 1.6(2) does not generalize to all polynomials, Solomon [6] showed that the displayed statement does indeed hold for *any* $f(x) \in K[x]$; indeed, the similarity is realized by a symmetric matrix. Solomon's proof (as well as one by Guralnick, also given in [6]) did not explicitly describe the similarity matrix S; we present here a proof in which such a symmetric matrix is constructed iteratively.

Theorem 2.1. *Let R be a ring with identity, and $f(x) \in R[x]$.*

If $f(x)$ is monic (not-necessarily-irreducible) in $R[x]$,
then the companion matrix A_f is similar to its transpose A_f^t.

More precisely, let $f(x) = x^n - \sum_{i=0}^{n-1} a_i x^i \in R[x]$ be an arbitrary monic polynomial of degree $n > 1$. Then A_f is similar to A_f^t, via the symmetric invertible matrix

$$S = \begin{pmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_{n-1} & 1 \\ -a_2 & -a_3 & -a_4 & \cdots & 1 & 0 \\ -a_3 & -a_4 & -a_5 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ -a_{n-1} & 1 & \cdots & 0 & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix};$$

specifically, $SA_f^t S^{-1} = A_f$. Moreover, S^{-1} is also a symmetric matrix.

of the upper triangular matrices. For small $n = 2, 3$, one can easily and directly establish this fact. For the general case, the centralizer of T_n is exactly the endomorphism ring of $V = K^n$ as a left T_n-module, so one has $\mathrm{End}({}_{T_n}V) = K$ because submodules of ${}_{T_n}V$ form a chain, and composition factors of ${}_{T_n}V$ are pairwise non-isomorphic. Therefore $M_n(K)$ is the double centralizer of T_n. The rings T_n ($n \in \mathbb{N}$) are the standard examples of serial rings, and ${}_{T_n}V$ is the longest uniserial T_n-module. For a piece of historical context regarding this idea, we mention that the now-ubiquitous notion of a *ring of operators* was defined by von Neumann via the double centralizer property.

Proof. For ease of notation let A denote A_f. Let α and α^t be the linear transformations defined by A and A^t with respect to the standard basis $\{\vec{e}_i\}$ on $V_R = R^n$. Put $\vec{f}_1 = \vec{e}_n = (A^t)^0 \vec{f}_1$, $\vec{f}_2 = A^t \vec{f}_1, \ldots, \vec{f}_n = (A^t)^{n-1} \vec{f}_1$. The proof becomes complete if we show that: $\{\vec{f}_i\}$ is a basis for V; the matrix of α^t with respect to the basis $\{\vec{f}_i\}$ is A; S is the transition matrix from $\{\vec{f}_i\}$ to $\{\vec{e}_i\}$; S is symmetric; and S^{-1} is also symmetric.

Since $A^t\vec{e}_2 = \vec{e}_1 + \vec{e}_n a_1, \ldots, A^t\vec{e}_{n-1} = \vec{e}_{n-2} + \vec{e}_n a_{n-2}$, $A^t\vec{e}_n = \vec{e}_{n-1} + \vec{e}_n a_{n-1}$, these vectors together with $\vec{e}_n = \vec{f}_1$ constitute a basis for V over R. The equality $\vec{e}_{n-1} = A^t\vec{e}_n - \vec{e}_n a_{n-1} = \vec{f}_2 - \vec{f}_1 a_{n-1}$ implies $A^t\vec{e}_{n-1} = A^t\vec{f}_2 - A^t\vec{f}_1 a_{n-1} = \vec{f}_3 - \vec{f}_2 a_{n-1} = \vec{e}_{n-2} + \vec{e}_n a_{n-2}$, whence $\vec{e}_{n-2} = \vec{f}_3 - \vec{f}_2 a_{n-1} - \vec{f}_1 a_{n-2}$. Continuing this process, an obvious induction shows

$$\vec{e}_{n-i} = \vec{f}_{i+1} - \vec{f}_i a_{n-1} - \cdots - \vec{f}_1 a_{n-i} = \vec{f}_{i+1} - \sum_{j=1}^{i} \vec{f}_{i-(j-1)} a_{n-j}$$

for $1 \le i \le n-1$. Therefore $\{\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_n\}$ is a basis of V_R, A is the matrix of α^t with respect to $\{\vec{f}_i\}$, and $S\vec{f}_i = \vec{e}_i$ holds for $i = 1, 2, \ldots, n$. Consequently, one has $SA^tS^{-1} = A$.

S is *left upper triangular* in the sense that $S_{i,j} = 0$ for all $j > n - i + 1$; additionally, the side diagonal of S is 1, i.e., $S_{i,n-i} = 1$ for all i.

For the symmetry of $S^{-1} = (\sigma_{i,j})$ one has to express each $\vec{f}_k$ as a linear combination of the $\{\vec{e}_i\}$. Simple induction shows that S^{-1} is *right lower triangular* in the sense that $\sigma_{i,j} = 0$ for all $j < n - i + 1$ and $\sigma_{i,n-1} = 1$ for all i. Furthermore, other entries are determined by

$$\sigma_{n-k+i,k+1} = \sigma_{n-k+i+1,k} \quad \text{for all } i < k, \quad \text{and} \quad \sigma_{k+1,n} = a_{n-k} + \sum_{i=1}^{k-1} a_{n-1}\sigma_{n-i+1,k}$$

with $\sigma_{n,1} = 1$ and $\sigma_{n-1,n} = a_{n-1}$. In particular, S^{-1} is symmetric. □

Example 2.2. If $f(x) = x^2 - 2 \in \mathbb{Q}[x]$ then as observed previously we have $A_f = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$, so that $A_f^t = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, and the symmetric matrix S constructed in Theorem 2.1 is $S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Clearly $S^{-1} = S$, and an easy computation verifies that $SA_f^tS^{-1} = A_f$.

Remark 2.3. We note that the matrix S constructed in Theorem 2.1 does not depend on a_0. We note also that Guralnick's proof [6, Appendix] gives (without computation) that any invertible similarity matrix S which transforms a companion matrix into its transpose must be symmetric.

It is well known that

any square matrix A over a (commutative) field K
is similar to its transpose A^t,

because the two matrices have the same set of invariant factors. Consequently, any square matrix A over a commutative domain is similar to its transpose A^t, if one allows the similarity matrix to take entries in the field of fractions of the domain. However, A and A^t need

not be similar over the domain itself; we give such an example, over $\mathbb{Z}$, in the Appendix below.

This observation raises the important (and apparently quite difficult) question of describing classes of rings R which have the property that any square matrix over R (or, any square matrix of a specified type over R) is similar to its transpose via a similarity matrix S having entries in R. Using Guralnick's module theoretic approach ([6, Appendix]) as a guide, we have been able to extend (Theorem 2.8) the previously displayed result from (commutative) fields to a significantly wider class of rings.

Definition 2.4. A ring T is called an *elementary divisor ring* in case every (finite) $m \times n$ matrix over T is diagonalizable over T. In other words, for all $m, n \in \mathbb{N}$ and each $m \times n$ matrix X with entries in T, there exist invertible $P \in M_m(T)$ and $Q \in M_n(T)$ with PXQ a diagonal matrix; that is, the (i, j) entry of PXQ is zero whenever $i \neq j$.

Remark 2.5. Finitely presented modules over elementary divisor rings are easily seen to be direct sums of finitely presented cyclic modules. This is a motivation for Kaplansky's invention of elementary divisor rings in [4], and his subsequent conjecture (solved in 1976) describing all commutative rings whose finitely generated modules are direct sums of cyclic ones. If R is a commutative von Neumann regular ring, then the polynomial ring $R[x]$ of a (central) variable x is an elementary divisor ring by [5]. Well-known examples of non-commutative elementary divisor rings include polynomial rings of one central variable over division rings, and even more generally, principal ideal domains (domains all of whose one-sided ideals are principal).[4]

Lemma 2.6. *Let R be any ring, $n \geq 1$, and $A \in M_n(R)$. Consider the free right R-module $V = R^n$, viewed as columns. Let $R[x]$ denote the usual polynomial ring in one variable with coefficients in R (so by definition, the symbol x commutes with the elements of R in $R[x]$). We define $* : V \times R[x] \to V$, as follows. For $f(x) = \sum_i x^i a_i$ with $a_i \in R$, and $v \in V$,*

$$v * f(x) = \sum_i A^i v a_i .$$

Then V is a right $R[x]$-module via $*$.

Proof. The linearity of $*$ over addition in V is clear. Now let $f(x) = \sum_i x^i a_i$ and $g(x) = \sum_j x^j b_j$ in $R[x]$. Then by definition (since x is central in $R[x]$) we have $f(x)g(x) =$

[4]Boolean rings (i.e., rings with $r^2 = r$ for all $r \in R$) and more generally, commutative algebras generated by (commuting) idempotents over fields, are prototypes for commutative von Neumann regular rings; these rings are precisely the rings of continuous functions from compact 0-dimensional Hausdorff spaces into fields (endowed with the discrete topology). For such R, one can see directly that $R[x]$ is an elementary divisor ring as follows. First, diagonalization is a local property on finitely many elements (the entries of the involved matrices); second, a subring generated by finitely many commuting idempotents is obviously a finite power of the base field after necessary orthogonalization; and third, polynomial rings over (skew) fields are elementary divisor rings. This argument, together with the trivial isomorphism $M_n(R[x]) \cong M_n(R)[x]$, provides even more classes of associative rings whose polynomial rings in one central variable are elementary divisor rings. These include locally semisimple rings (i.e., unions of semisimple artinian rings), and so, in particular, ultramatricial algebras over a field K. (These are by definition the countable direct union of subalgebras, each of which is a finite direct sum of finitely many finite-dimensional matrix algebras over K.) Ultramatricial algebras play an important role in the classification of certain operator algebras via K-theory.

$h(x) = \sum_k x^k c_k$, where $c_k = \sum a_i b_{k-i}$. We must show $v*(f(x)g(x)) = (v*f(x))*g(x)$. We have

$$v * (f(x)g(x)) = v * h(x) = \sum\nolimits_k A^k v c_k.$$

On the other hand,

$$\begin{aligned}(v * f(x)) * g(x) &= \left(\sum\nolimits_i A^i v a_i\right) * g(x) = \sum\nolimits_j A^j \left(\sum\nolimits_i A^i v a_i\right) b_j \\ &= \sum\nolimits_{i,j} A^j A^i (v a_i b_j) = \sum\nolimits_k A^k v \left(\sum\nolimits_{i+j=k} a_i b_j\right) = \sum\nolimits_k A^k v c_k,\end{aligned}$$

as desired. □

Remark 2.7. We note that the right action of $R[x]$ given in Lemma 2.6 was described (in abbreviated form) by Guralnick in [6, Appendix]. This action is somewhat subtle, and apparently has not been widely utilized in the literature.

Indeed, the intuition behind this right action seems to become more transparent if we step away from the particular situation where $V = R^n$, and consider an arbitrary right R-module W_R. So in this more general context there is no way of realizing the action of R as endomorphisms on W_R (unlike the situation in R^n, where we have available the left R-action $r(r_1, \ldots, r_n)^t = (rr_1, \ldots, rr_n)^t$). But, every endomorphism A on W_R (written on the left) induces the right module action of $R[x]$ on W_R by setting, for each $w \in W$ and $f(x) = \sum_i x^i a_i \in R[x]$,

$$w * f(x) = \sum A^i w a_i.$$

The main point is that for describing a right $R[x]$-module structure on W_R where x is central, it is enough to define an action of x on W_R which is exactly an endomorphism A of W_R, without any need to give an interpretation of any possible meaning of a ring extension of R by A.

We are now in position to give the previously mentioned generalization.

Theorem 2.8. *Let R be a ring for which the polynomial ring $R[x]$ (with central variable x) is an elementary divisor ring. Then*

any square matrix A over R is similar to its transpose A^t.

Proof. By Lemma 2.6 we have that $V = R^n$ is a right $R[x]$-module with the indicated action. By [6, Theorem A.2] (which is a generalization to the noncommutative case of a result in Buccino's 1967 University of Chicago Ph.D. thesis), $V_{R[x]}$ is isomorphic to $R[x]^n/M$, where M is the right $R[x]$-submodule of $R[x]^n$ generated by $\{vx - Av \mid v \in V = R^n \subseteq R[x]^n\}$. By the hypothesis on $R[x]$, there exist invertible matrices $B, C \in M_n(R[x])$ such that $B(x - A)C$ is a diagonal matrix Δ with entries $\lambda_i \in R[x]$ $(1 \leq i \leq n)$. Consequently, $V_{R[x]}$ is isomorphic to $\oplus_{i=1}^n R[x]/\lambda_i R[x]$ as right $R[x]$-modules. Since $\Delta = \Delta^t = (B(x - A)C)^t = C^t(x - A^t)B^t$, the canonical right $R[x]$-modules induced by A and A^t (respectively) on $R_R^n = V$ are isomorphic. But then invoking [6, Theorem A.1] yields the similarity of A and A^t. □

3 The algebraic closure of a field

The standard proof of the existence and uniqueness of an algebraic closure of a field K up to K-isomorphism is an immediate and nice application of Zorn's Lemma. However, it is the authors' experience that students often struggle mightily to understand this proof, in part because the resulting maximal algebraic extension of K seems to simply fall out of the sky, with no real concrete context provided for what the elements of the algebraic closure look like, or where they live.

We present in Theorem 3.1 what we believe to be a significantly more concrete construction of the algebraic closure of a field. Although we (not surprisingly) invoke Zorn's Lemma, we remain consistent with the thrust of this article by again using matrices as context. However, unlike the matrices considered previously, the matrices we consider here are infinite-dimensional. We first give some context.

Let V be an n-dimensional vector space over a field K for some $n \in \mathbb{N}$. Of course it is well known that any linear transformation T from V to V (i.e., a *linear operator on* V) can be described by an $n \times n$ matrix M. Depending on how one chooses to write functions (interpreting $f \circ g$ as either "first g, then f", or vice-versa), the ith column (resp., row) of M is $T(\vec{e}_i)$, the image under T of the ith basis vector. Since here we have chosen to consider vector spaces with scalar action on the right, we choose the column interpretation for matrices. More formally, we have $\mathrm{End}(V_K) \cong M_n(K)$ via this association. This dual interpretation of operators, both as matrices and as their coordinate-free description as linear transformations (endomorphisms), often helps to make proofs of various results simpler and more transparent, and this context is no exception.

It is not hard to show directly that this same behavior extends to linear operators on infinite-dimensional vector spaces as well. If the dimension of V is cardinality $\aleph$, one considers (imagines?) the collection of $\aleph \times \aleph$ matrices with entries in K. Since any element of V is by definition a (finite) linear combination of basis vectors $\{\vec{e}_i \mid i \in I\}$, then any operator T on V must take each basis vector $\vec{e}_i$ to some element of V of the form $k_1\vec{e}_{i_1} + \cdots + k_\ell\vec{e}_{i_\ell}$. So (depending on what notational choice you made in the previous paragraph), the ring of all linear operators on V can be viewed as the ring $CFM_\aleph(K)$ (resp., $RFM_\aleph(K)$), the $\aleph \times \aleph$ matrices where each column (resp., each row) contains at most finitely many nonzero elements of K. The association in the infinite case is the same as in the finite-dimensional setting: the ith column (resp., row) of M is $T(\vec{e}_i)$, the image under T of the ith basis vector. For the same reason as mentioned previously, here we choose to consider $CFM_\aleph(K)$, and we have $\mathrm{End}(V_K) \cong CFM_\aleph(K)$ via this association.

As in the finite-sized matrix case, we can still clearly view K itself as being embedded in either $CFM_\aleph(K)$ or $RFM_\aleph(K)$ as the scalar matrices. More generally, for any $n \in \mathbb{N}$, we can view $M_n(K)$ as being embedded in $CFM_\aleph(K)$ or $RFM_\aleph(K)$, by associating each $M \in M_n(K)$ with the $n \times n$ block-diagonal matrix which has every block equal to M.

We are now in position to present the aforementioned "concrete" construction of the algebraic closure of a field K.

Theorem 3.1. *Let V be an infinite dimensional vector space over a field K such that the dimension of V is either:*

– *uncountable, when K is finite, or*

– *larger than the cardinality of K, when K is infinite.*

Then all maximal algebraic field extensions of K in $\mathrm{End}(V_K)$ *are algebraically closed, and hence are algebraic closures of K.*

In addition, any two such maximal algebraic field extensions are conjugate, via a suitable inner automorphism of $\mathrm{End}(V_K)$.

Proof. Let F be a subfield of $\mathrm{End}(V_K)$ which is a maximal algebraic field extension of K. Such subfields exist by a relatively straightforward Zorn Lemma argument. Then V can be viewed as a vector space over F. The condition imposed on both the dimension of V_K and the cardinality of K implies that V is also infinite dimensional over F.

If F is not algebraically closed, then there is an irreducible polynomial $p(x) \in F[x]$ having $\deg(p(x)) > 1$, whence the companion matrix A_p of $p(x)$ is a root of $p(x)$ by Proposition 1.1. So by the previously-described embedding of A_p into $\mathrm{End}(V_F) \subseteq \mathrm{End}(V_K) \cong CFM_{\aleph}(K)$ one obtains $\mu \in CFM_{\aleph}(K)$ such that the subalgebra generated by F and μ is a proper algebraic field extension of F, a contradiction to the maximality of F. Therefore every maximal algebraic K-subalgebra of $\mathrm{End}(V_K)$ is algebraically closed.

If E is another maximal algebraic K-algebra subfield of $\mathrm{End}(V_K)$, then a second relatively easy application of Zorn's Lemma yields a K-isomorphism $\phi\colon E \to F$ providing the standard proof for the uniqueness of algebraic closures of fields up to isomorphism. Subsequently, V can be considered as a vector space in two different ways: by multiplication via $A \in E$, or by multiplication via $\phi(A)$, respectively. Since the dimension of V_K is either bigger than $|K|$ when K is infinite or uncountable when K is finite, V has the same dimension as a vector space over both E and F. Consequently, a straightforward extension of the argument already used in Proposition 1.6(2) shows that E, F are necessarily conjugates. □

By using the same ideas as presented in the previous proof, we get the following somewhat more general result.

Proposition 3.2. *Let V be an infinite-dimensional vector space over a field K, and let F be a subfield of* $\mathrm{End}(V_K)$ *maximal* (*with respect to inclusion*) *in the set of field extensions of K contained in* $\mathrm{End}(V_K)$. *If V is infinite-dimensional over F, then F is algebraically closed. Moreover, any two such maximal subfields E, F are conjugate via an element of* $\mathrm{Aut}(V_K)$, *provided that V has the same* (*not necessarily infinite*) *dimension as a vector space over E and F, and the corresponding vector spaces V_E, V_F are isomorphic.*

For context vis-à-vis the cardinality conditions imposed in the two preceding results, there may exist fields K and vector spaces V_K and maximal K-subalgebras E in $\mathrm{End}(V_K)$ which are fields, such that V_E is finite-dimensional. Examples of this type can be constructed using Lüroth's theorem on endomorphisms of fields of rational functions which show that such maximal subfields need not be algebraically closed.

4 Further applications of the companion matrix

In this final section we make passing mention of some additional seemingly disparate places in which companion matrices may play a clarifying role.

4.1 Galois fields

Galois' construction of finite fields is mathematically quite marvelous, but often difficult for students to grasp, in part because it is somewhat opaque as to how the addition should work. The use of companion matrices in the construction of finite fields is well known to experts, and can be found in most monographs on finite fields, but as with other topics is not particular popular in introductory textbooks.

Companion matrices identify finite fields as subfields of matrices over prime fields $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. By [3, Part I. Fields, Theorem 52], $x^p - x - au^{p-1}$ is irreducible over $K(u)$ if $\mathrm{char}(K) = p$, $x^p - x - a$ is irreducible over K, and u is a root of $x^p - x - a$. Therefore, starting from $\mathbb{F}_p$ and the irreducible polynomial $x^p - x - 1$ over $\mathbb{F}_p$, one can easily construct Galois fields $\mathbb{F}_{p^{p^n}}$ of cardinality p^{p^n} $(n > 0)$ as subalgebras generated by the associated companion matrices in the corresponding matrix rings of degree p over the previously built ones.

For $\mathbb{F}_{p^n}$ $(n > 1)$ one needs to guarantee the existence of irreducible polynomials of degree n over $\mathbb{F}_p$; this can be obtained by counting irreducible polynomials iteratively in increasing degree, similar to a sieve method for locating prime numbers. It is worth noting that irreducible polynomials over $\mathbb{F}_p$ are prime factors modulo p of cyclotomic polynomials, and the latter are independent of the characteristic of the field, even irreducible over the field $\mathbb{Q}$ of rationals. (We do not know how to achieve this result by using companion matrices.)

In any case we can identify Galois fields as follows.

Theorem 4.1. *For any prime number p and a positive integer n all maximal subfields F of the matrix ring $M_n(\mathbb{F}_p)$ over the Galois field $\mathbb{F}_p$ have p^n elements satisfying the double centralizer property. Consequently, all maximal subfields of $M_n(\mathbb{F}_p)$ are conjugate via appropriate elements of* $\mathrm{GL}_n(\mathbb{F}_p)$.

Proof. Without loss of generality one can assume $n > 1$. We consider $M_n(\mathbb{F}_p)$ as the endomorphism ring of a vector space V of dimension n over $\mathbb{F}_p$ and assume that F has p^k elements. Then k is a divisor of n. Put $n = kl$. Then V is a vector space of dimension l over F in the obvious way. As we already remarked, there exists an irreducible polynomial f of degree l over F and therefore the companion matrix A of f over F is a linear transformation on V, whence an element in $M_n(\mathbb{F}_p)$. Consequently, the F-subalgebra E of $\mathrm{End}(V_F)$ is a field extension of F contained in $M_n(\mathbb{F}_p)$. This shows $l = 1$, $k = n$ and hence F has p^n elements and the double centralizer property. The last claim is now obvious. □

Since the Galois field $\mathbb{F}_{p^n}$ consists of 0 together with $p^n - 1$ $(p^n - 1)$st roots of unity, every Galois field $\mathbb{F}_{p^k}$ can be identified with the unique subfield of $\mathbb{F}_{p^\ell}$ $(k|\ell)$ consisting of all $(p^k - 1)$st roots of unity together with 0. This shows that the set of all finite extensions of

the Galois field $\mathbb{F}_p$ form a directed set under inclusion via the above described embedding and therefore their directed union (direct limit) is algebraically closed. Consequently we obtain the following obvious but not well-noted examples of algebraically closed fields.

Theorem 4.2. *For each prime number p the set of all finite extensions of the Galois field $\mathbb{F}_p$ is a directed set under inclusion whose union is the algebraic closure $\bar{\mathbb{F}}_p$ of $\mathbb{F}_p$. Roughly speaking, the field extension of $\mathbb{F}_p$ consisting of all roots of unity is the algebraic closure $\bar{\mathbb{F}}_p$ of $\mathbb{F}_p$.*

Theorem 4.2 shows that one cannot omit the cardinality condition in either Theorem 3.1 or Theorem 3.2. Namely, the usual cardinality argument implies that the vector space $V_{\mathbb{F}_p}$ of countably infinite dimension over the Galois field $\mathbb{F}_p$ can be considered as a vector space of any countable dimension over the algebraic closure of $\mathbb{F}_p$. Consequently, in the endomorphism ring $\mathrm{End}(V_{\mathbb{F}_p})$ there are two subfields K_1 and K_2 which are algebraic closures of $\mathbb{F}_p$ such that the dimensions of V over K_1 or K_2, respectively, are different. Therefore although K_1 and K_2 are $\mathbb{F}_p$-isomorphic, they are not conjugate via an element of $\mathrm{Aut}(V_{\mathbb{F}_p})$, that is, any such isomorphism is not extendable to an inner isomorphism of $\mathrm{End}(V_{F_p})$. This example together with Theorem 3.1 and Theorem 3.2 reflect some aspects of and are closely related to the previously-mentioned Noether–Skolem theorem, see for example, [3, Part II. Rings, Theorem 51].

Moreover, the multiplicative group $\bar{\mathbb{F}}_p^* = \bar{\mathbb{F}}_p \setminus \{0\}$ of $\bar{\mathbb{F}}_p$ shows also the beauty and ingenuity of Galois' construction. Since equations $x^l = a$ $(\in \bar{\mathbb{F}}_p^*)$ are solvable, $\bar{\mathbb{F}}_p^*$ is an abelian divisible torsion group, in particular, a direct sum of q-subgroups of all (q^l)th roots of unity which are isomorphic to quasi-cyclic groups $C(q^\infty)$, where q runs over all prime divisors of $p^k - 1 (k \in \mathbb{N})$. This shows clearly a big difference between the additive and multiplicative groups of $\bar{\mathbb{F}}_p$, respectively. While the former is the vector space over $\mathbb{F}_p$, the latter is a direct summand of the additive group $\mathbb{Q}/\mathbb{Z}$. However, we do not know, for example, precisely which primes appear in the prime factorization of $p^n - 1$ $(n \in \mathbb{N})$, even in the case $p = 2$. This last comment is directly related to a difficult number theory problem: for example, the Mersenne primes appear naturally in finding the prime factorization of $2^p - 1$ where p runs over odd primes.

In summary, the argument shows immediately that the socle (i.e., the sum of minimal multiplicative subgroups) of a field is a sum of cyclic groups of prime order with multiplicity 1. Thus the multiplicative subgroup of all roots of unity in an algebraically closed field is either $\mathbb{Q}/\mathbb{Z}$ or a proper direct summand of $\mathbb{Q}/\mathbb{Z}$, according to whether the characteristic of the field is 0 or not. Since any finite subgroup of $\mathbb{Q}/\mathbb{Z}$ is cyclic, we obtain the following well-known result together with an explanation.

Corollary 4.3. *The group of all roots of unity in an algebraically closed field is either $\mathbb{Q}/\mathbb{Z}$ or a direct summand of $\mathbb{Q}/\mathbb{Z}$, according to whether the characteristic of the field is 0 or not. In particular, any finite subgroup of the multiplicative group of a field, which of necessity must be a subgroup of the roots of unity of the field, is cyclic.*

It is perhaps less well known that the above result no longer holds for division rings. For example, the quaternion group $\{\pm 1, \pm i, \pm j, \pm k\}$ in the division ring $\mathbb{H}$ of quaternions is not cyclic.

4.2 Transcendence

Matrices shed a light on transcendence. By Cantor's ingenious cardinality argument, almost all irrational real numbers are transcendental; but verifying the transcendence of a particular real number is quite hard, one reason being the existence of algebraic numbers which are not rational. In contrast, every $n \times n$ matrix over a field K is algebraic over K (by the Cayley–Hamilton theorem). Therefore, in the matrix setting, one must look to infinite-sized matrices to find transcendental matrices.

More formally, let A be an element of $CFM_{\aleph}(K)$ which is algebraic over K; so A is a root of a polynomial $f(x) = p_i^{l_1}(x) \cdots p_k^{l_k}(x)$, where $p_1(x), \ldots, p_k(x)$ are irreducible polynomials in $K[x]$. Then in the usual way, $V = K^{(\aleph)}$ can be considered as a module over $R = K[x]/f(x)K[x] \cong \oplus_{i=1}^{k} K[x]/p_i^{l_i}(x)K[x] = \oplus_{i=1}^{k} R_i$, a finite direct sum of uniserial Artinian rings R_i. Therefore ${}_RV = \oplus_{i=1}^{k} {}_{R_i}V_i$ and by [7, Theorem 6.7] each V_i is a finite direct sum of modules V_{i_j} which are direct sums of isomorphic uniserial Artinian R_i-modules of length $\leq l_i$. Consequently, with respect to an appropriate basis of V_{i_j} the restriction of A to V_{i_j} is either a finite matrix with a fixed companion matrix appearing repeatedly on the diagonal or a *periodic* matrix, that is, an infinite matrix with a fixed companion matrix appearing on the diagonal repeatedly depending on the cardinality of the K-dimension of V_{i_j}. If we call finite direct sums (i.e., finite block sums) of such matrices *ultimately periodic matrices*, then A is algebraic if and only if with respect to an appropriate basis, it is ultimately periodic; equivalently, A is similar to an ultimately periodic matrix. This gives not only a normal form of algebraic matrices but also gives a description of transcendental infinite column-finite matrices. Moreover, there are no algebraic irrational infinite matrices, in contrast to the fact that there are algebraic irrational numbers. Therefore we have verified

Theorem 4.4. *An infinite column-finite matrix is either transcendental or algebraic. It is algebraic if and only if it is similar to an ultimately periodic matrix.*

4.3 Algebraic curves

Last, we briefly note that algebraic curves are parameterized by companion matrices. As an example, let $P(x, y) = \sum_{i=0}^{n} y^i p_{n-i}(x)$ be a(n irreducible) polynomial in two variables x, y, written as a polynomial in y with coefficients $p_i(x) \in K[x]$. The companion matrix of $\bar{P}(t, y) = p_0^{-1}(t)P(t, y)$ $(t \in K)$ parameterizes $P(x, y)$, where singularities appear when either $p_0(t) = 0$ or $\bar{P}(t, y) \in K[y]$ is reducible. Moreover, the companion matrix of $\bar{P}(x, y)$ over the rational function field $K(x)$ identifies an algebraic function y as a matrix solution to $P(x, y) = 0$ in term of x over the rational function field $K(x)$.

5 Appendix
A matrix over $\mathbb{Z}$ which is not similar over $\mathbb{Z}$ to its transpose

We present here the details of an example of the type mentioned subsequent to Remark 2.3. The point is this: if K is a field, then every $M \in M_n(K)$ is similar to its transpose. If R is a commutative domain, then R embeds in its field of fractions; so in particular every $N \in M_n(R)$ is similar to its transpose, viewed as a matrix in $M_n(K)$. However, it need not be the case that such N is similar to its transpose using a similarity matrix from $M_n(R)$.

We show this behavior for the specific matrix $M = \begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix} \in M_2(\mathbb{Z})$. To wit, we show that M is not similar to its transpose via a similarity matrix in $M_2(\mathbb{Z})$. For consider the system of Diophantine equations resulting from a proposed matrix equation

$$\begin{pmatrix} 0 & 2 \\ 3 & 1 \end{pmatrix} = T \begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix} T^{-1}$$

with $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z})$. The invertibility of T in $M_2(\mathbb{Z})$ yields that $ad - bc = \delta \in \{1, -1\}$ (in particular, $\delta^2 = 1$). By contradiction assume $T^{-1} = \begin{pmatrix} u & v \\ x & y \end{pmatrix} \in M_2(\mathbb{Z})$.

Then solving u, v, x, y in terms of a, b, c, d via Cramer's rule we have

$$u = \frac{\delta(3a+b)}{3}, \quad v = \frac{\delta(3a+b-6c-2d)}{6}, \quad x = -\delta b, \quad y = \frac{\delta(2d-b)}{3}.$$

Therefore substituting u, x in the equalities $au + bx = 1$, $cu + dx = 0$ guaranteed by $TT^{-1} = 1$ shows

$$\delta a(3a+b) = 1 + \delta b^2, \quad c(3a+b) = bd.$$

Multiplying the first one on both sides by c, substituting the second into the resulting equation, and using $\delta^2 = 1$ and $ad = bc + \delta$ gives

$$\delta ac(3a+b) = \delta abd = \delta b(bc+\delta) = \delta b^2 c + b.$$

But $\delta ac(3a+b) = c(1+\delta b^2) = c + c\delta b^2$, so that $\delta b^2 c + b = c + c\delta b^2$, so that $b = c$. Moreover substituting v, y in $cu + dy = 0$ given by $TT^{-1} = 1$ and using $c(3a+b) = bd, b = c$ we have

$$\frac{a(3a+b-6c-2d}{6} + \frac{b(2d-b)}{3} = 0,$$

so that

$$3a^2 - 5ac - 2ad + 4bd - 2b^2 = 3a^2 - 5ac + 4c(3a+b) - 2ad + 2bc = 0,$$

which gives in turn that

$$3a^2 + 7ac = 2\delta = a(3a+7c).$$

The last two equalities imply that both a and c are odd, and also that a is a divisor of 2, whence $a \in \{1, -1\}$, and so $\pm(7c \pm 3) = 2\delta$, which is impossible. □

References

[1] M. Eichler, *Introduction to the theory of algebraic numbers and functions*. Academic Press, 1966.

[2] J. Fraleigh, *A First Course in Abstract Algebra, 7*th *Ed.* Addison Wesley, 2003.

[3] I. Kaplansky, *Fields and Rings*. University of Chicago Press, 1969.

[4] I. Kaplansky, Elementary divisors and modules, *Trans. AMS* **66** (1949), 464–491.

[5] Th.S. Shores, Modules over semihereditary Bezout rings, *Proc. AMS* **46** (1974), 211–213.

[6] L. Solomon, Similarity of the companion matrix and its transpose with an appendix by Robert M. Guralnick, *Linear Algebra Appl.* **302/303** (1999), 555–561.

[7] D.W. Sharpe and P. Vámos, *Injective modules*. Cambridge University Press, 1972.

Gene Abrams
Department of Mathematics
University of Colorado
Colorado Springs, CO 80918, USA
e-mail: abrams@math.uccs.edu

P.N. Ánh[5]
Rényi Institute of Mathematics
Hungarian Academy of Sciences
H-1364 Budapest, Pf. 127 Hungary
e-mail: anh.pham.ngoc@renyi.mta.hu

[5]The second author was partially supported by National Research, Development and Innovation Office NKHIH K119934, Vietnamese Institute for Advanced Study in Mathematics (VIASM), and Vietnamese Institute of Mathematics, VAST