Where to look for the zeros of a polynomial

Stefan Veldsman*

Abstract

The well-known Kronecker construction of an extension of a ring containing a zero for a given monic polynomial over the ring, is usually given as a quotient ring of a ring of polynomials. This quotient ring can also be seen as a matrix ring. In this note, a description of this matrix ring is given.

One of the early joys of a first undergraduate course in ring theory (or abstract algebra), is the result that any monic polynomial over a ring with unity will always have a zero in the coefficient ring or in an extension thereof. This well-known result of Kronecker is usually shown by embedding the coefficient ring into the quotient ring of the polynomial ring over the ideal generated by the polynomial. As such, the elements are cosets which can also be written as polynomials of formal sums, all with degree less than the degree of the starting polynomial. Products of such formal sums give powers of the indeterminate which must be reduced to the required degree by using a rule prescribed by the starting polynomial.

More than often, the beauty and significance of this process and result is lost on the students at this early stage of their training in algebra. This is mainly due to their discomfort of working with rings where the elements are cosets or formal sums.

There is a more natural way to view this ring extension which requires a minimum number of tools. Most introductory texts will refer to this only in the exercises (if at all) by considering one or two special cases. The particular case that is usually mentioned is the one that leads to the circulant matrices.

Here we will present the general case. This approach, using matrices, is not new in the sense that bits and pieces of the ideas involved have appeared in different contexts elsewhere. However, the simplicity and the minimal requirements of the method does warrant wider exposure. A further advantage of this approach is that many properties of the extension ring can be expressed in terms of well-known linear algebra concepts.

We start by illustrating the idea with an easy example before we outline the general approach. If $\mathbb{Z}[x]$ denotes the ring of polynomials over the ring of integers

Received 3 May 2008; accepted for publication 22 July 2008.
*Department of Mathematics and Statistics, Sultan Qaboos University, Muscat, Sultanate of Oman. E-mail: veldsman@squ.edu.om

$\mathbb{Z}$, let $h(x) = x^3 + 2x - 1 \in \mathbb{Z}[x]$ and let $\mathbb{M}_3(\mathbb{Z}, h)$ be the set of 3×3 matrices defined by:

$$\mathbb{M}_3(\mathbb{Z}, h) = \left\{ \begin{bmatrix} a & b & c \\ c & a-2c & b \\ b & c-2b & a-2c \end{bmatrix} \;\middle|\; a, b, c \in \mathbb{Z} \right\}.$$

This set of matrices, which has been chosen in a very specific way to be outlined below, is a subring of the ring $\mathbb{M}_3(\mathbb{Z})$ of all 3×3 matrices over $\mathbb{Z}$. The ring $\mathbb{Z}$ can be embedded into $\mathbb{M}_3(\mathbb{Z}, h)$ by

$$a \mapsto \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}.$$

The polynomial $h(x) = x^3 + 2x - 1$ over $\mathbb{Z}$ can thus be regarded as a polynomial over $\mathbb{M}_3(\mathbb{Z}, h)$ and it has a zero in $\mathbb{M}_3(\mathbb{Z}, h)$ namely

$$t := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -2 & 0 \end{bmatrix}.$$

Indeed,

$$\begin{aligned} h(t) &= t^3 + 2t - 1 \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -2 & 0 \end{bmatrix}^3 + \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -2 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= 0. \end{aligned}$$

The matrix

$$t = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -2 & 0 \end{bmatrix}$$

is the well-known companion matrix of the polynomial $h(x)$ and plays an important role in the development of the theory below.

We start by recalling the Kronecker construction of the required ring extension. This is followed by introducing suitable notation to describe the multiplication in this ring and to define the matrices required in the ring extension. The necessary computational rules are established to show that the Kronecker construction is isomorphic to this ring of matrices.

Let A be a commutative ring with identity 1. Let

$$h(x) := x^k - p_{k-1}x^{k-1} - \cdots - p_1 x - p_0 \in A[x]$$

be a monic polynomial of degree k over A, $k \geq 2$. The commutativity of A is assumed for convenience of exposition; most of the initial results will be valid under the weaker assumption that the coefficients $p_0, p_1, \ldots, p_{k-1}$ of $h(x)$ are in

the centre of A. Let H denote the ideal in $A[x]$ generated by $h(x)$. As is well known, A can be embedded into $A[x]/H$ and

$$\frac{A[x]}{H} \cong \left\{ a_1 + a_2 y + a_3 y^2 + \cdots + a_k y^{k-1} \mid a_i \in A, y^k = \sum_{i=0}^{k-1} p_i y^i \right\},$$

where $a + H$ is identified with $a \in A$ and $x + H$ with y. The polynomial $h(x)$ over A can be identified with $\overline{h}$ as a polynomial over $A[x]/H$ and $\overline{h}(y) = 0$. The set $\{1 = y^0, y, y^2, \ldots, y^{k-1}\}$ is linearly independent over A. Addition in $A[x]/H$ is straightforward and the multiplication is as usual for polynomials except that all powers $y^{k+l}, l \geq 0$, must be reduced to linear combinations of $y^0, y, y^2, \ldots, y^{k-1}$ using $y^k = \sum_{i=0}^{k-1} p_i y^i$. To give the general rule for this multiplication, we fix some notation.

For i and j nonnegative integers, define elements $e(i, j)$ of A inductively by:

$$e(i,j) := \begin{cases} 1 & \text{if } i = j = 0, \\ 0 & \text{if } i = 0 \text{ or } j = 0 \text{ but not both}, \\ e(i-1, j-1) + p_{j-1} e(i-1, k) & \text{for } 1 \leq j \leq k \text{ and } i = 1, 2, 3, \ldots. \end{cases}$$

For later use, note that

$$e(i,j) = \begin{cases} 1 & \text{if } 1 \leq i = j \leq k, \\ 0 & \text{if } 1 \leq j \leq k, i \neq j \end{cases}$$

and $e(k+1, j) = p_{j-1}$ for $j = 1, 2, 3, \ldots, k$.

Our first observation, which can easily be verified by induction on s, is:

Property 1. For $i \geq 1$, $1 \leq j \leq k$ and $1 \leq s \leq \min\{i, j\}$,

$$\begin{aligned} e(i,j) &= e(i-s, j-s) + \sum_{l=1}^{s} e(k+1, j-l) e(i-l, k) \\ &= e(i-s, j-s) + \sum_{l=1}^{s} p_{j-l} e(i-l, k). \end{aligned}$$

The main tool used in describing the multiplication in $A[x]/H$ is:

Property 2. For any $s \geq 1$, $y^s = \sum_{j=1}^{k} e(s+1, j) y^{j-1}$. In particular, for $l \geq 0$, $y^{k+l} = \sum_{j=1}^{k} e(k+l+1, j) y^{j-1}$ and if $0 \leq i \leq k-1$, the coefficient of y^i in y^{k+l} is $e(k+l+1, i+1)$.

Indeed, for $s = 1$, it is straightforward to check that $\sum_{j=1}^{k} e(2,j)y^{j-1} = y$. Suppose thus $s \geq 1$ and that the statement is true for s. Then

$$\begin{aligned}
y^{s+1} &= y^s y \\
&= \sum_{j=1}^{k} e(s+1,j)y^j \text{ by the induction assumption} \\
&= \sum_{j=1}^{k-1} e(s+1,j)y^j + e(s+1,k)\sum_{j=1}^{k} p_{j-1}y^{j-1} \text{ since } y^k = \sum_{j=0}^{k-1} p_j y^j \\
&= \sum_{j=2}^{k} (e(s+1,j-1) + p_{j-1}e(s+1,k))y^{j-1} + (e(s+1,0) + p_0 e(s+1,k))y^0 \\
&= \sum_{j=1}^{k} (e(s+1,j-1) + p_{j-1}e(s+1,k))y^{j-1} \\
&= \sum_{j=1}^{k} e(s+2,j)y^{j-1} \text{ as required.}
\end{aligned}$$

The product of two elements from $A[x]/H$ is then given:

Property 3.

$$\begin{aligned}
&(a_1 + a_2 y + a_3 y^2 + \cdots + a_k y^{k-1})(b_1 + b_2 y + b_3 y^2 + \cdots + b_k y^{k-1}) \\
&\qquad = \sum_{i=1}^{k} \left(\sum_{l=1}^{k} a_l \left(b_{i+1-l} + \sum_{s=2}^{l} e(k+s-1,i) b_{k-l+s} \right) \right) y^{i-1}.
\end{aligned}$$

To verify this statement, note that for $0 \leq i \leq k-1$ the coefficient of y^i in the product above is:

$$\begin{aligned}
&a_1 b_{i+1} + a_2(b_i + \text{ coefficient of } y^i \text{ in } b_k y^k) \\
&\quad + a_3(b_{i-1} + \text{ coefficient of } y^i \text{ in } (b_{k-1}y^k + b_k y^{k+1})) \\
&\quad + \cdots + a_r(b_{i-r+2} + \text{ coefficient of } y^i \\
&\qquad\qquad \text{in } (b_{k-r+2}y^k + b_{k-r+3}y^{k+1} + \cdots + b_k y^{k+r-2})) \\
&\quad + \cdots + a_k(b_{i-k+2} + \text{ coefficient of } y^i \text{ in } (b_2 y^k + b_3 y^{k+1} + \cdots + b_k y^{k+(k-2)})) \\
&= a_1 b_{i+1} + a_2(b_i + e(k+1,i+1)b_k) + a_3(b_{i-1} + e(k+1,i+1)b_{k-1} \\
&\quad + e(k+2,i+1)b_k) \\
&\quad + \cdots + a_r(b_{i-r+2} + e(k+1,i+1)b_{k-r+2} + e(k+2,i+1)b_{k-r+3} \\
&\qquad\qquad + \cdots + e(k+r-1,i+1)b_k) \\
&\quad + \cdots + a_k(b_{i-k+2} + e(k+1,i+1)b_2 + e(k+2,i+1)b_3 \\
&\qquad\qquad + \cdots + e(k+(k-1),i+1)b_k) \\
&= \sum_{l=1}^{k} a_l \left(b_{i-l+2} + \sum_{s=2}^{l} e(k+s-1,i+1) b_{k-l+s} \right).
\end{aligned}$$

Hence the result follows (making the necessary adjustments in the final statement for the coefficient of y^{i-1} rather than that of y^i as determined above).

In order to write the elements $a_1+a_2y+a_3y^2+\cdots+a_ky^{k-1}$ of $A[x]/H$ as matrices, each y^{i-1} will be replaced by a $k\times k$ matrix $E_i, i=1,2,3,\ldots,k$. These matrices will be described by using the row matrices e_r defined below. We do not distinguish between $1\times k$ row matrices and k-dimensional vectors from $A^k = A\oplus A\oplus\cdots\oplus A$ and use whatever is more convenient in a particular situation. For each $r\geq 1$, let

$$e_r := \begin{bmatrix} e(r,1) & e(r,2) & \cdots & e(r,k)\end{bmatrix}.$$

Then, for example,

$$e_1 = [1\ 0\ 0\ \cdots\ 0],\qquad e_2 = [0\ 1\ 0\ \cdots\ 0],\ldots,\qquad e_k = [0\ 0\ \cdots\ 0\ 1],$$
$$e_{k+1} = [p_0\ p_1\ p_2\ \cdots\ p_{k-1}]$$

and

$$e_{k+2} = [p_0p_{k-1}\ \ p_0+p_1p_{k-1}\ \ p_0+p_2p_{k-1}\ \ \cdots\ \ p_0+p_{k-1}p_{k-1}].$$

Sometimes e_{k+1} is denoted by p. At times we will also need the $k\times 1$ column vectors

$$c(r,j) := \begin{bmatrix} e(r,j)\\ e(r+1,j)\\ \vdots\\ e(r+k-1,j)\end{bmatrix}\qquad \text{for } r\geq 1 \text{ and } 1\leq j\leq k.$$

For $r\geq 1$, let E_r be the $k\times k$ matrix with ith row e_{r+i-1} (or jth column $c(r,j)$). This means

$$E_r = \begin{bmatrix} e_r\\ e_{r+1}\\ \vdots\\ e_{r+k-1}\end{bmatrix}$$
$$= \begin{bmatrix} e(r,1) & e(r,2) & \ldots & e(r,k)\\ e(r+1,1) & e(r+1,2) & \ldots & e(r+1,k)\\ \vdots & \vdots & \ddots & \vdots\\ e(r+k-1,1) & e(r+k-1,2) & \ldots & e(r+k-1,k)\end{bmatrix}$$
$$= \begin{bmatrix} c(r,1) & c(r,2) & \cdots & c(r,k)\end{bmatrix},$$

where the (i,j)th entry is given by $e(r+i-1,j)$. In particular, $E_1 = I_k$, the $k\times k$ identity matrix and

$$E_2 = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0\\ 0 & 0 & 1 & \ldots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \ldots & 1\\ p_0 & p_1 & p_2 & \cdots & p_{k-1}\end{bmatrix}$$

is the companion matrix of the polynomial $h(x) = x^k - p_{k-1}x^{k-1} - \cdots - p_1x - p_0$. This matrix has been around for a long time; amongst others it has been shown

that the Cayley–Hamilton theorem is valid and so $h(x) = \mathrm{char}(E_2) = \det(xI_k - E_2)$ where char denotes the characteristic polynomial of E_2 and det denotes the determinant. Moreover, $h(x)$ is the minimal polynomial of E_2. The entries of the matrix $E_r = [e(r+i-1, j)]_{k\times k}$ satisfy the recursion

$$e(r+i-1,j) = e(r+i-2,j-1) + p_{j-1}e(r+i-2,k) \qquad \text{for all } i,j = 1,2,3,\ldots,k.$$

This recursion can also be expressed in terms of the rows and columns. Let $\sigma\colon A^k \to A^k$ denote the shift function $\sigma([a_1 \;\; a_2 \;\; \cdots \;\; a_k]) := [0 \;\; a_1 \;\; a_2 \;\; \cdots \;\; a_{k-1}]$. For example, $\sigma(e_r) = e_{r+1}$ for $r = 1,2,3,\ldots,k-1$ and $\sigma(e_k) = 0$. Note that σ is an A-module homomorphism (both left and right) and for any $r \geq 1$, $e_{r+1} = \sigma(e_r) + e(r,k)p$ where p denotes the row vector $p := e_{k+1} = [p_0 \;\; p_1 \;\; p_2 \;\; \cdots \;\; p_{k-1}]$. Moreover,

$$\begin{aligned} e_{r+1} &= \sum_{l=1}^{k} e(r+1,l)e_l \\ &= \sum_{l=1}^{k} (e(r,l-1) + p_{l-1}e(r,k))e_l \\ &= \sum_{l=1}^{k} e(r,l-1)e_l + e(r,k)p \end{aligned}$$

and it follows that $\sigma(e_r) = \sum_{l=1}^{k} e(r,l-1)e_l$. In terms of the columns, it can be shown that $c(r,j) = c(r-1,j-1) + p_{j-1}c(r-1,k)$ for $r \geq 2$ and $1 \leq j \leq k$. The products of the matrices E_r are described in:

Property 4. For any $r,s \geq 1$, $E_rE_s = E_{r+s-1}$. In particular, this means $E_rE_s = E_sE_r$ and $E_2^i = E_{i+1}$ for all $i = 1,2,3,\ldots$.

For $r = 1$ or $s = 1$, the statement is clearly true; suppose thus $r \geq 2$ and proceed by induction on s. For $s = 2$, $E_rE_2 = [a_{ij}]$, say, where $a_{ij} = e(r+i-1) \bullet c(2,j)^t$. Here $\bullet$ denotes the dot product and $(\cdot)^t$ denotes the transpose. Thus

$$\begin{aligned} a_{ij} &= \sum_{l=1}^{k} e(r+i-1,l)e(2+l-1,j) \\ &= e(r+i-1,j-1) + p_{j-1}e(r+i-1,k) \\ &= e(r+i,j) \\ &= e(r+1+i-1,j) \end{aligned}$$

since $1 \leq l \leq k, 1 \leq j \leq k$ and $e(l+1,j) = 0$ unless $l+1 = j$ or $l = k$. But this is just the (i,j)th entry of E_{r+1} which shows $E_rE_2 = E_{r+1}$ for all $r \geq 2$. Suppose the result holds for all $r \geq 2$ and some $s \geq 2$. Then $E_rE_{s+1} = E_r(E_sE_2) = (E_rE_s)E_2 = (E_{r+s-1})E_2 = E_{(r+s-1)+1} = E_{(r+(s+1))-1}$ as required.

Our next objective is to show that the matrices E_r can be expressed as a linear combination of $E_1, E_2, \ldots, E_k$ and we start with:

Property 5.

$$E_{k+1} = \sum_{l=1}^{k} p_{l-1} E_l.$$

Firstly, using induction on i, it can be shown that

$$e(k+i, j) = \sum_{l=1}^{i} e(k+1, j-i+l) e(k+l-1, k)$$

for $1 \leq i \leq k$ and $i \leq j \leq k$. Secondly, from the first part

$$\begin{aligned}\sum_{l=1}^{k} e(k+1, l) e(i+l-1, k) &= \sum_{l=k-i+1}^{k} e(k+1, l) e(i+l-1, k) \\ &= \sum_{l=1}^{i} e(k+1, k-i+l) e(k-1+l-1, k) \\ &= e(k+i, k)\end{aligned}$$

since $e(i+l-1, k) = 0$ for $i+l-1 \leq k-1$. We now show that $E_{k+1} = \sum_{l=1}^{k} e(k+1, l) E_l$ by comparing the rows of the two matrices. The first row of $\sum_{l=1}^{k} e(k+1, l) E_l$ is

$$\begin{aligned}&e(k+1,1)e_1 + e(k+1,2)e_2 + \cdots + e(k+1,k)e_k \\ &\quad = \begin{bmatrix} e(k+1,1) & e(k+1,2) & \cdots & e(k+1,k) \end{bmatrix}\end{aligned}$$

which is also the first row of E_{k+1}. Let $1 \leq i < k$ and suppose the ith row of E_{k+1} coincides with the ith row of $\sum_{l=1}^{k} e(k+1, l) E_l$, that is

$$e_{k+i} = e(k+1,1)e_i + e(k+1,2)e_{i+1} + \cdots + e(k+1,k)e_{i+k-1}.$$

The $(i+1)$th row of E_{k+1} is $e_{k+i+1} = \sigma(e_{k+i}) + e(k+i, k)p$. On the other hand, the $(i+1)$th row of $\sum_{l=1}^{k} e(k+1, l) E_l$ is:

$$\begin{aligned}&e(k+1,1)e_{i+1} + e(k+1,2)e_{i+2} + \cdots + e(k+1,k-i)e_{i+(k-i)} \\ &\quad + e(k+1,k-i+1)e_{i+(k-i)+1} + \cdots + e(k+1,k)e_{k+i} \\ &\qquad = e(k+1,1)(\sigma(e_i) + e(i,k)p) + e(k+1,2)(\sigma(e_{i+1}) + e(i+1,k)p) + \cdots \\ &\qquad\quad + e(k+1,k-i)(\sigma(e_{k-1}) + e(k-1,k)p) + e(k+1,k-i+1) \\ &\qquad\quad \times (\sigma(e_k) + e(k,k)p) + \cdots + e(k+1,k)(\sigma(e_{k+i-1}) + e(k+i-1,k)p) \\ &\qquad = \sigma(e(k+1,1)e_i + e(k+1,2)e_{i+1} + \cdots + e(k+1,k-i)e_{k-1} \\ &\qquad\quad + e(k+1,k-i+1)e_k + \cdots + e(k+1,k)e_{k+i-1}) + (e(k+1,1)e(i,k) \\ &\qquad\quad + e(k+1,2)e(i+1,k) + \cdots + e(k+1,k-i)e(k-1,k) \\ &\qquad\quad + e(k+1,k-i+1)e(k,k) + \cdots + e(k+1,k)e(k+i-1,k))p \\ &\qquad = \sigma(e_{k+i}) + \left(\sum_{l=1}^{k} e(k+1,l)e(i+l-1,k)\right)p \\ &\qquad = \sigma(e_{k+i}) + e(k+i,k)p \\ &\qquad = e_{k+i+1}\end{aligned}$$

as required.

Property 6. For any $r \geq 1$, $E_r = \sum_{l=1}^{k} e(r,l)E_l = \sum_{l=1}^{k} e(r,l)E_2^{l-1}$.

The first equality can be verified by using (5) above and induction on r and the second equality follows from (4).

Let $\mathbb{M}_k(A)$ be the ring of all $k \times k$ matrices over A and let $\mathbb{M}_k(A,h) := \{a_1E_1 + a_2E_2 + \cdots + a_kE_k \mid a_i \in A\}$. In view of properties (4) and (5) above, $\mathbb{M}_k(A,h)$ is a subring of $\mathbb{M}_k(A)$. If we define $E_2^0 := E_1$ (which is just I_k) and recall that $E_2^i = E_{i+1}$, we can write $\mathbb{M}_k(A,h) = \{a_1E_2^0 + a_2E_2 + a_3E_2^2 + \cdots + a_kE_2^{k-1} \mid a_i \in A\}$ and think of $\mathbb{M}_k(A,h)$ as the subring of $\mathbb{M}_k(A)$ generated by $\{E_2\} \cup \{aE_1 \mid a \in A\}$. This is sometimes written as $A[E_2] = \{f(E_2) \mid f(x) \in A[x]\}$. Even though $\mathbb{M}_k(A)$ is not commutative, $\mathbb{M}_k(A,h)$ is commutative. The powers of E_2 are reduced to linear combinations of $E_1, E_2, \ldots, E_k$ using (4) and (5) above. We want to determine the product of two elements in $\mathbb{M}_k(A,h)$ explicitly. Let $\sum_{i=1}^{k} a_iE_i$, $\sum_{i=1}^{k} b_iE_i \in \mathbb{M}_k(A,h)$. For $1 \leq r \leq k$ the coefficient of E_r in the product $(\sum_{i=1}^{k} a_iE_i)(\sum_{i=1}^{k} b_iE_i)$ is given by:

$$\begin{aligned}
&a_1b_r + a_2(b_{r-1} + \text{ coefficient of } E_r \text{ in } b_kE_{k+1}) \\
&\quad + a_3(b_{r-2} + \text{ coefficient of } E_r \text{ in } (b_{k-1}E_{k+1} + b_kE_{k+2})) + \cdots \\
&\quad + a_l(b_{r-l+1} + \text{ coefficient of } E_r \\
&\qquad \text{in } (b_{k-l+2}E_{k+1} + b_{k-l+3}E_{k+2} + \cdots + b_kE_{k+l-1})) \\
&\quad + \cdots + a_k(b_{r-k+1} + \text{ coefficient of } E_r \\
&\qquad \text{in } (b_2E_{k+1} + b_3E_{k+2} + \cdots + b_kE_{k+k-1})) \\
&\quad = \sum_{l=1}^{k} a_l(b_{r-l+1} + \sum_{s=2}^{l} b_{k-l+s}e(k+s-1,r)).
\end{aligned}$$

Thus

$$\left(\sum_{i=1}^{k} a_iE_i\right)\left(\sum_{i=1}^{k} b_iE_i\right) = \sum_{r=1}^{k}\left(\sum_{l=1}^{k} a_l\left(b_{r-l+1} + \sum_{s=2}^{l} b_{k-l+s}e(k+s-1,r)\right)\right)E_r.$$

From the preceding, we thus have

Theorem 1. *For the unital commutative ring A and $h(x) := x^k - p_{k-1}x^{k-1} - \cdots - p_1x - p_0 \in A[x]$, the quotient ring*

$$A[x]/H = \left\{a_1 + a_2y + a_3y^2 + \cdots + a_ky^{k-1} \;\middle|\; a_i \in A, y^k = \sum_{i=0}^{k-1} p_iy^i\right\}$$

is isomorphic to $\mathbb{M}_k(A,h)$.

The ring A has the canonical embedding into $\mathbb{M}_k(A,h)$. This ring of matrices contains a zero for $h(x)$ regarded as a polynomial over $\mathbb{M}_k(A,h)$, namely its companion matrix E_2. Indeed, $h(E_2) = E_2^k - \sum_{l=1}^{k} p_{l-1}E_2^{l-1} = E_{k+1} - \sum_{l=1}^{k} p_{l-1}E_2^{l-1} = 0$ (by (5)). Moreover, any zero for $h(x)$ which is already present in A remains a zero for $h(x)$ in $\mathbb{M}_k(A,h)$. By iterating this process, one can obtain as many different zeros for $h(x)$ in a suitable matrix extension of A as is desired (see the examples below).

One of the main advantages of working with matrices from $\mathbb{M}_k(A,h)$ rather than formal sums $a_1 + a_2y + a_3y^2 + \cdots + a_ky^{k-1}$, is that no reductions are necessary when calculating products. The price to pay for this is that the elements of the extension ring containing the zeros for the polynomial are $k \times k$ matrices rather than just formal sums or k-dimensional vectors. The examples below include many of the well-known special matrix rings.

Examples

(a) Let $h(x) = x^2 + 1 \in \mathbb{R}[x]$. Then

$$\mathbb{M}_2(\mathbb{R},h) = \left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \;\middle|\; a,b \in \mathbb{R} \right\} \cong \mathbb{C},$$

where $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex numbers respectively. The polynomial $h(x)$ has two zeros in $\mathbb{M}_2(\mathbb{R},h)$, namely

$$E_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad -E_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Let $\mathbb{C}_1 := \mathbb{M}_2(\mathbb{R},h)$. Repeating this procedure, $h(x)$ has at least four zeros in $\mathbb{C}_2 := \mathbb{M}_2(\mathbb{C}_1,h)$ $(\subseteq \mathbb{M}_4(\mathbb{R}))$, namely

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(the embeddings of $E_2 = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ and $-E_2 = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ into $\mathbb{M}_2(\mathbb{C}_1,h)$) as well as

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

(the new E_2 and $-E_2$ in $\mathbb{M}_2(\mathbb{C}_1,h)$). Then $\mathbb{C}_3 := \mathbb{M}_2(\mathbb{C}_2,h)$ $(\subseteq \mathbb{M}_{2^3}(\mathbb{R}))$ will have at least six zeros for $h(x)$ regarded as a polynomial over $\mathbb{C}_3$. In general, $\mathbb{C}_{n+1} := \mathbb{M}_2(\mathbb{C}_n,h)$ will have at least $2(n+1)$ zeros for $h(x)$.

The general form of the polynomial above is $h(x) = x^k + 1 \in \mathbb{R}[x]$ which leads to the $k \times k$ skew-circulant matrix ring

$$\mathbb{M}_k(\mathbb{R},h) = \left\{ \begin{bmatrix} a_1 & a_2 & \dots & a_k \\ -a_k & a_1 & \dots & a_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ -a_2 & -a_3 & \dots & a_1 \end{bmatrix} \;\middle|\; a_i \in \mathbb{R} \right\}.$$

(b) Let $h(x) = x^k - 1 \in \mathbb{R}[x]$. Then

$$\mathbb{M}_k(\mathbb{R}, h) = \left\{ \begin{bmatrix} a_1 & a_2 & \dots & a_k \\ a_k & a_1 & \dots & a_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_2 & a_3 & \dots & a_1 \end{bmatrix} \;\middle|\; a_i \in \mathbb{R} \right\}$$

is the $k \times k$ circulant matrix.

(c) Let $h(x) = x^k \in \mathbb{R}[x]$. Then

$$\mathbb{M}_k(\mathbb{R}, h) = \left\{ \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_k \\ 0 & a_1 & a_2 & \dots & a_{k-1} \\ 0 & 0 & a_1 & \dots & a_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_1 \end{bmatrix} \;\middle|\; a_i \in \mathbb{R} \right\}.$$

(d) Let $h(x) = x^2 + x + 1 \in \mathbb{Z}_2[x]$. Then

$$\mathbb{M}_3(\mathbb{Z}_2, h) = \left\{ \begin{bmatrix} a & b \\ b & a+b \end{bmatrix} \;\middle|\; a, b \in \mathbb{Z}_2 \right\}$$

is the well-known four-element field.

We conclude with a few remarks to highlight the role of results from linear algebra in the properties of the ring $\mathbb{M}_k(A, h)$.

Property 7. For any given $h(x) = x^k - p_{k-1}x^{k-1} - \cdots - p_1 x - p_0 \in A[x]$ of degree k and $r \geq 1$,

$$\det(E_r) = \begin{cases} p_0^{r-1} & \text{if } k \text{ is odd,} \\ (-1)^{r-1} p_0^{r-1} & \text{if } k \text{ is even.} \end{cases}$$

Once again, an induction on r provides justification for this statement. If p_0 is a unit, then E_2^{-1} exists and

$$E_2^{-1} = \begin{bmatrix} \frac{-p_1}{p_0} & \frac{-p_2}{p_0} & \dots & \frac{-p_{k-1}}{p_0} & \frac{1}{p_0} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \in \mathbb{M}_k(A, h).$$

Note that by (4) above, $E_{i+1}^{-1} = (E_2^{-1})^i$.

Property 8. As mentioned earlier, the Cayley–Hamilton theorem is valid which means for $h(x) \in A[x]$ we have $h(x) = \mathrm{char}(E_2) = \det(xI_k - E_2)$ and $h(E_2) = 0$.

Property 9. For $M \in \mathbb{M}_k(A, h)$, M will be a unit (i.e. it is invertible) if and only if $\det(M)$ is a unit of A. In such a case, $M^{-1} = g(M)$ for some $g(x) \in A[x]$ (cf. Corollary 7.25 of Brown [1]) which means $M^{-1} \in \mathbb{M}_k(A, h)$. We conclude by describing $\det(M)$ for $M \in \mathbb{M}_k(A, h)$ where A is an integral domain. Suppose $h(x) = x^k - p_{k-1}x^{k-1} - \cdots - p_1 x - p_0 \in A[x]$ has a factorisation $h(x) =$

$(x-\omega_1)(x-\omega_2)\dots(x-\omega_k)$ for some $\omega_i \in A$. Suppose $M = \sum_{i=1}^{k} a_i E_2^{i-1}$. Then $f(x) := a_1 + a_2 x + \cdots + a_k x^{k-1} \in A[x]$ and $f(E_2) = M$. We show $\det(M) = f(\omega_1)f(\omega_2)\dots f(\omega_k)$. Indeed, using Theorems 8.54 and 8.50 of Brown [1], it follows that $\det(M) = \mathcal{R}(h,f) = f(\omega_1)f(\omega_2)\dots f(\omega_k)$ where $\mathcal{R}(h,f)$ denotes the resultant of $h(x)$ and $f(x)$.

A special case worth mentioning is the following. For each $i \geq 0$, let

$$W(i) := \begin{bmatrix} \omega_1^i & \omega_2^i & \dots & \omega_k^i \\ \omega_1^{i+1} & \omega_2^{i+1} & \dots & \omega_k^{i+1} \\ \omega_1^{i+2} & \omega_2^{i+2} & \dots & \omega_k^{i+2} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1^{i+k-1} & \omega_2^{i+k-1} & \dots & \omega_k^{i+k-1} \end{bmatrix}.$$

Let $D(a_1, a_2, \dots, a_k)$ be the $k \times k$ diagonal matrix with a_i in position (i,i). For any $j \geq 0$, $W(j+1) = W(j)D(\omega_1, \omega_2, \dots, \omega_k)$ and so $W(j) = W(0)D(\omega_1^j, \omega_2^j, \dots, \omega_k^j)$. Recall that

$$W(0) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \omega_1 & \omega_2 & \dots & \omega_k \\ \omega_1^2 & \omega_2^2 & \dots & \omega_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1^{k-1} & \omega_2^{k-1} & \dots & \omega_k^{k-1} \end{bmatrix}$$

is the well-known Vandermonde matrix with $\det(W(0)) = \prod_{i>j}(\omega_i - \omega_j)$. Since $h(\omega_i) = 0$, $\omega_i^k = \sum_{j=1}^{k} p_{j-1}\omega^{j-1}$ and so $E_2 W(0) = W(1)$. Hence $E_2^j W(0) = W(j) = W(0)D(\omega_1^j, \omega_2^j, \dots, \omega_k^j)$ for any $j \geq 0$. Then

$$MW(0) = W(0)D(f(\omega_1), f(\omega_2), \dots, f(\omega_k)).$$

If $\det(W(0))$ is a unit, then

$$(W(0))^{-1} M W(0) = D(f(\omega_1), f(\omega_2), \dots, f(\omega_k)).$$

M has eigenvalues $f(\omega_1), f(\omega_2), \dots, f(\omega_k)$ with corresponding eigenvectors $(1, \omega_1, \omega_1^2, \dots, \omega_1^{k-1})$, $(1, \omega_2, \omega_2^2, \dots, \omega_2^{k-1}), \dots, (1, \omega_k, \omega_k^2, \dots, \omega_k^{k-1})$ respectively.

References

[1] Brown, C. (1993). *Matrices Over Commutative Rings*. Marcel Dekker, New York.