Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

ON THE REDUCTION OF MATRIX POLYNOMIALS TO HESSENBERG FORM*

THOMAS R. CAMERON†

Abstract. It is well known that every real or complex square matrix is unitarily similar to an upper Hessenberg matrix. The purpose of this paper is to provide a constructive proof of the result that every square matrix polynomial can be reduced to an upper Hessenberg matrix, whose entries are rational functions and in special cases polynomials. It will be shown that the determinant is preserved under this transformation, and both the finite and infinite eigenvalues of the original matrix polynomial can be obtained from the upper Hessenberg matrix.

Key words. Matrix polynomial, Polynomial eigenvalue problem, Arnoldi, Krylov subspace, Self-adjoint, Invariant subspace.

AMS subject classifications. 15A15, 15A18, 15A21, 15A54.

1. Introduction.

Recently, there has been a push to develop methods for finding the eigenvalues and eigenvectors of a matrix polynomial that are similar to the methods that have been used on real and complex matrices. For example, in [5], the authors develop Arnoldi type methods that operate on the coefficient matrices, and effectively project a large problem onto a smaller problem. In [7, 8] it is shown that any matrix polynomial can be reduced to triangular or quasi-triangular form, while preserving the degree and the finite and infinite elementary divisors of the matrix polynomial; the constructive proof requires information on the finite and infinite elementary divisors of the original matrix polynomial. Starting with the elementary divisors of a matrix polynomial is not practical if our end goal is to solve for the eigenvalues of the matrix polynomial. However, in [8], the authors also discuss using structure preserving similarities acting on the linearization of the matrix polynomial, in order to obtain a triangular form of the matrix polynomial. This is truly exciting research, a reliable process for reducing a matrix polynomial to triangular form while preserving the eigenvalues would have obvious benefits.

In this paper, we will show that every matrix polynomial can be reduced to an upper Hessenberg matrix whose entries are rational functions and in special cases

*Received by the editors on May 17, 2015. Accepted for publication on April 20, 2016. Handling Editor: James G. Nagy.
†Department of Mathematics, Washington State University, Pullman, Washington (tcameron@math.wsu.edu).

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

polynomials. Recall that every matrix polynomial $P(z)$ admits the representation

$$P(z) = E(z)D(z)F(z),$$

where $D(z)$ is a diagonal matrix polynomial whose diagonal entries are the invariant polynomials of $P(z)$; $E(z)$ and $F(z)$ are matrix polynomials with constant nonzero determinants. This is known as the Smith form of the matrix polynomial $P(z)$ [4]. This Smith form is attained through elementary row and column operations which are stored in the matrix polynomials $E(z)$ and $F(z)$. With this result in mind, it is not surprising that we can reduce a matrix polynomial to upper Hessenberg form.

Our approach to obtaining an upper Hessenberg form of a matrix polynomial is unique, in that we choose to think of a matrix polynomial as a linear operator from $\mathbb{F}^n$ to $\mathbb{F}^n$, where $\mathbb{F}$ is the field of rational functions. In Section 2, we define a pseudo inner product on $\mathbb{F}^n \times \mathbb{F}^n$, and R-similarity of matrices in $\mathbb{F}^{n\times n}$. Armed with this pseudo inner product, we are able to apply Arnoldi and Krylov subspace methods to matrix polynomials. For a concise review on these methods, see [9]. In Section 3, we use these familiar techniques to provide an elegant and constructive proof of the result that every square matrix polynomial is similar to an upper Hessenberg matrix A in $\mathbb{F}^{n\times n}$. We note that these results can be generalized for non-square matrix polynomials and matrix rationals. The practical importance of this reduction is that using Hyman's method [3, 10] and the Ehrlich-Aberth method [1, 2, 6], we can develop a numerical method for computing the eigenvalues of A. The theoretical importance is that in proving this result we will show that many familiar techniques from linear algebra can be applied directly to matrix polynomials, without using a linearization, as long as we work over the field of rational functions. In Section 4, we show that, in general, R-similarity transformations in $\mathbb{F}^{n\times n}$ do not preserve the Smith form of the original matrix polynomial. After identifying the problem, we provide sufficient conditions for which the Smith form is preserved under such transformations.

2. The field of rational functions. A matrix polynomial of size $n \times n$ and degree d is defined by

$$P(z) = \sum_{k=0}^{d} A_k z^k, \tag{2.1}$$

where $A_k \in \mathbb{C}^{n\times n}$, $z \in \mathbb{C}$, and $A_d \neq 0$. Let $\mathbb{F}$ denote the field of rational functions over $\mathbb{C}$ and $\mathbb{F}^n$ denote the vector space of all n-tuples whose entries are rational functions with complex coefficients. It will be useful to denote the matrix polynomial as $P = [p_{ij}]_{i,j=1}^n$, where p_{ij} is a scalar polynomial in z, whose k-th coefficient is given by the (i, j)-th entry of the k-th coefficient matrix in (2.1). A matrix polynomial can be considered as a linear transformation $P : \mathbb{F}^n \to \mathbb{F}^n$. In this sense, the matrix polynomial in (2.1) has been written in the standard basis. The definition of linear

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

independence and a spanning set of vectors in $\mathbb{F}^n$ is the same as it would be in $\mathbb{C}^n$, except our scalars are rational functions. We are interested in representing a matrix polynomial P with respect to other bases for $\mathbb{F}^n$, with this in mind we note the following definition.

Definition 2.1. We say that two matrices $A, B \in \mathbb{F}^{n\times n}$ are R-similar, if there exists a matrix $S \in \mathbb{F}^{n\times n}$, such that $\det S$ is a nonzero rational function and $AS = SB$.

Let $S \in \mathbb{F}^{n\times n}$ and consider the elementary transformations of interchanging two rows, and adding to some row another row multiplied by a rational function. With this in mind, it is not hard to show that the matrix S has linearly independent columns if and only if $\det S$ is a nonzero rational function.

In Section 3, we will use the Arnoldi method to show that every matrix polynomial is similar to an upper Hessenberg matrix A in $\mathbb{F}^{n\times n}$.

Definition 2.2. Define the pseudo inner product $\langle\cdot, \cdot\rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ by

$$\langle u, v\rangle = u_1 \cdot \overline{v_1} + \cdots + u_n \cdot \overline{v_n},$$

where $(\cdot)$ and $(+)$ are multiplication and addition of rational functions. The complex conjugate is taken over the coefficients of v_i, for $i = 1, \ldots, n$.

The function in Definition 2.2 is not a traditional inner product, since the output is a rational function. However, the following theorem will show that the pseudo inner product satisfies linearity in the first argument, and $\langle u, u\rangle = 0$ if and only if $u = 0$.

Theorem 2.3. *Let* $u, v, z \in \mathbb{F}^n$ *and* $c \in \mathbb{F}$. *Then*

$$\begin{aligned}\langle cu, v\rangle &= c\langle u, v\rangle, \\ \langle u+v, z\rangle &= \langle u, z\rangle + \langle v, z\rangle, \\ \langle u, u\rangle = 0 \quad &\textit{if and only if} \ \ u = 0.\end{aligned}$$

Proof. Since $\mathbb{F}^n$ is a vector space over the field of rational functions F we have

$$\langle cu, v\rangle = cu_1\bar{v}_1 + \cdots + cu_n\bar{v}_n = c\langle u, v\rangle,$$

and

$$\begin{aligned}\langle u+v, z\rangle &= (u_1+v_1)\cdot\bar{z}_1 + \cdots + (u_n+v_n)\cdot\bar{z}_n \\ &= (u_1\cdot\bar{z}_1 + \cdots + u_n\cdot\bar{z}_n) + (v_1\cdot\bar{z}_1 + \cdots + v_n\cdot\bar{z}_n) = \langle u, z\rangle + \langle v, z\rangle.\end{aligned}$$

If $u = 0$, then it is clear that $\langle u, u\rangle = 0$. To prove the converse, first note that each element of u can be expressed as $u_i = \frac{p_i}{q_i}$, where p_i is a polynomial of degree d_i. Let

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

$d = \max_{1\le i\le n} d_i$, then write $p_i = a_{i,d}z^d + \cdots + a_{i,1}z + a_{i,0}$. Therefore, u can be written as

$$u = \begin{bmatrix} \frac{a_{1,d}z^d}{q_1} \\ \vdots \\ \frac{a_{n,d}z^d}{q_n} \end{bmatrix} + \cdots + \begin{bmatrix} \frac{a_{1,0}}{q_1} \\ \vdots \\ \frac{a_{n,0}}{q_n} \end{bmatrix} := v_d + \cdots + v_0.$$

By linearity of the inner product,

$$\langle u,\, u\rangle = \langle v_0,\, v_0\rangle + \langle v_0,\, v_1\rangle + \cdots + \langle v_d,\, v_d\rangle\,,$$

where

$$\langle v_i,\, v_i\rangle = z^{2i}\left(\frac{|a_{1,i}|^2}{|q_1|^2} + \cdots + \frac{|a_{n,i}|^2}{|q_n|^2}\right).$$

If $\langle u,\, u\rangle = 0$, then the highest degree term $\langle v_d,\, v_d\rangle = 0$. But this implies that $\max_{1\le i\le n} d_i < d$, and we can apply the same argument to v_i for $i = d-1,\, \ldots,\, 0$. Therefore, $\langle v_i,\, v_i\rangle = 0$ for $i = 0,\, 1,\, \ldots,\, d$ and it follows that $u = 0$. ∎

In addition to the properties in Theorem 2.3, the pseudo inner product satisfies conjugate symmetry, if the complex conjugate is taken over the coefficients of the rational function. We proceed to define orthogonality in $\mathbb{F}^n$ with respect to our pseudo inner product.

Definition 2.4. Two vectors $u,\, v \in \mathbb{F}^n$ are orthogonal, if $\langle u,\, v\rangle = 0$.

The properties that we have proven our inner product possesses are all we need to obtain upper Hessenberg form. This is because the Arnoldi process which we will employ is truly the Gram-Schmidt process in disguise. Consider the following example.

Example 2.5. Let $v_1(z) = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $v_2(z) = \begin{bmatrix} iz \\ 1 \end{bmatrix}$. With the inner product in Definition 2.2, we can use the Gram-Schmidt process to find an orthogonal basis for $\mathbb{F}^2$. Let $u_1(z) = v_1(z)$ and

$$u_2(z) = v_2(z) - \frac{\langle v_2,\, u_1\rangle}{\langle u_1,\, u_1\rangle}u_1(z) = \begin{bmatrix} iz \\ 1 \end{bmatrix} - \frac{i(z-1)}{2}\begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} \frac{i}{2}(z+1) \\ \frac{1}{2}(z+1) \end{bmatrix}.$$

One can easily verify that $\langle u_1,\, u_2\rangle = 0$.

In general, we may have a set of linearly independent vectors $\{v_1,\, \ldots,\, v_n\}$ in $\mathbb{F}^n$. Let $u_1 = v_1$ and proceed to define

$$u_{k+1} = v_{k+1} - \sum_{i=1}^{k} \frac{\langle v_{k+1},\, u_i\rangle}{\langle u_i,\, u_i\rangle}u_i,$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

for $k = 1, \ldots, n-1$. Suppose we have created a set of orthogonal vectors $\{u_1, \ldots, u_k\}$ in $\mathbb{F}^n$. By linearity of our inner product

$$\langle u_{k+1}, u_j \rangle = \langle v_{k+1}, u_j \rangle - \sum_{i=1}^{k} \frac{\langle v_{k+1}, u_i \rangle}{\langle u_i, u_i \rangle} \langle u_i, u_j \rangle ,$$

for $j = 1, \ldots, k$. Since $\langle u_i, u_j \rangle = 0$ for all $i \neq j$, it follows that $\langle u_{k+1}, u_j \rangle = 0$ for $j = 1, \ldots, k$. Therefore, with the properties proved in Theorem 2.3, we can be sure that the Gram-Schmidt process under our inner product will produce orthogonal vectors in $\mathbb{F}^n$.

3. Arnoldi and Krylov subspace methods. In Section 2, we saw that we could build an orthogonal basis for $\mathbb{F}^n$ by using n linearly independent vectors in $\mathbb{F}^n$ and the Gram-Schmidt process. In this section, we will give a practical method for finding an orthogonal basis for a Krylov subspace of $\mathbb{F}^n$.

3.1. The Arnoldi method. Let P be an $n \times n$ matrix polynomial, and let $v \in \mathbb{F}^n$ be a nonzero vector. Then the sequence of vectors

$$v,\, Pv,\, P^2v,\, \ldots \tag{3.1}$$

is known as a Krylov sequence generated by the vector v and matrix polynomial P. Since the vector space $\mathbb{F}^n$ is an $n-$dimensional space over the field $\mathbb{F}$, it follows that the above Krylov sequence can produce at most n linearly independent vectors. Suppose that the vectors

$$v,\, Pv,\, \ldots,\, P^{k-1}v, \quad k \le n$$

are linearly independent. Then these vectors form a basis for the $k-$dimensional Krylov subspace

$$K_k(P, v) = \operatorname{span}\left\{v,\, Pv,\, \ldots,\, P^{k-1}v\right\}.$$

We are after an orthogonal basis for the Krylov subspace $K_k(P, v)$. To this end, consider the Arnoldi method which starts with the nonzero vector $v_1 = v$ and then produces

$$\hat{v}_{j+1} = Pv_j - \sum_{i=1}^{j} \frac{\langle Pv_j, v_i \rangle}{\langle v_i, v_i \rangle} v_i \quad \text{and} \quad v_{j+1} = \frac{1}{\beta_{j+1}} \hat{v}_{j+1} \tag{3.2}$$

for $j = 1, \ldots, k-1$. We will refer to $\beta_{j+1} \in \mathbb{F}$ as a scaling factor. We have some freedom in how we choose our scaling factor, but for now we let $\beta_{j+1} = 1$. We will consider other scaling factors and how they affect our transformation in Section 4.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

Note that the vectors that (3.2) produces are proportional, with respect to our scaling factor, to the vectors we would obtain from applying the Gram-Schmidt process to the vectors $v,\, Pv,\, \ldots,\, P^{k-1}v$. Therefore,

$$\operatorname{span}\left\{v,\, Pv,\, \ldots,\, P^{j}v\right\} = \operatorname{span}\left\{v_1,\, v_2,\, \ldots,\, v_{j+1}\right\},$$

for $j = 1,\, \ldots,\, k-1$.

The Arnoldi process combined with our inner product has provided us a way of computing an orthogonal basis for the Krylov subspace $K_k(P,\, v)$. Let k be the largest number of linearly independent vectors that the Krylov sequence in (3.1) can produce. If $k = n$, then the Arnoldi process will produce an orthogonal basis for $\mathbb{F}^n$, and the matrix representation of P with respect to this basis will be upper Hessenberg.

Example 3.1. Let $P(z) = \begin{bmatrix} z^2 & 1 & z \\ 1 & z^2 & 1 \\ z & 1 & z^2 \end{bmatrix}$ and $v_1(z) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Applying the Arnoldi process, with scalar factor 1, we obtain

$$v_2(z) = P(z)v_1(z) - z^2 v_1(z) = \begin{bmatrix} 0 \\ 1 \\ z \end{bmatrix},$$

$$v_3(z) = P(z)v_2(z) - \frac{z\left(z^3+z+2\right)}{z^2+1}v_2(z) - \left(z^2+1\right)v_1(z) = \begin{bmatrix} 0 \\ \frac{z\left(z^2-1\right)}{z^2+1} \\ \frac{1-z^2}{z^2+1} \end{bmatrix},$$

$$v_4(z) = P(z)v_3(z) - \frac{z\left(z^3+z-2\right)}{z^2+1}v_3(z) - \frac{\left(z^2-1\right)^2}{\left(z^2+1\right)^2}v_2(z) - 0v_1(z) = 0.$$

Therefore, we have

$$V(z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{z\left(z^2-1\right)}{z^2+1} \\ 0 & z & \frac{1-z^2}{z^2+1} \end{bmatrix} \quad \text{and} \quad A(z) = \begin{bmatrix} z^2 & z^2+1 & 0 \\ 1 & \frac{z\left(z^3+z+2\right)}{z^2+1} & \frac{\left(z^2-1\right)^2}{\left(z^2+1\right)^2} \\ 0 & 1 & \frac{z\left(z^3+z-2\right)}{z^2+1} \end{bmatrix},$$

such that $P(z)V(z) = V(z)A(z)$.

The matrix $V(z)$ in Example 3.1 has orthogonal columns, and it follows from Definition 2.1 that $P(z)$ is R-similar to $A(z)$. Therefore, the fact that $\det P(z) = \det A(z)$ is not surprising. We generalize this result in the following theorem.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

THEOREM 3.2. *Suppose that* $P(z)$, $A(z) \in \mathbb{F}^{n\times n}$ *are R-similar. Then* $\det\left(z^\alpha P(z^\beta)\right) = \det\left(z^\alpha A(z^\beta)\right)$ *for all* α, $\beta \in \mathbb{Z}$.

Proof. If $P(z)$ is R-similar to $A(z)$, then there exists a $V(z) \in \mathbb{F}^{n\times n}$ with linearly independent columns such that P(z)V(z)=V(z)A(z). Therefore, $\det V(z)$ is a nonzero rational function, and it follows that $\det P(z) = \det A(z)$. Moreover, by multiplicativity of the determinant, $\det\left(z^\alpha P(z^\beta)\right) = \det\left(z^\alpha A(z^\beta)\right)$. ◻

The finite eigenvalues of $P(z)$ are the roots of $\det P(z)$. Therefore, by Theorem 3.2 that the finite eigenvalues of $P(z)$ and $A(z)$ are equal. The infinite eigenvalues of $P(z)$ are defined as the zero valued roots of $\det\left(z^d P(z^{-1})\right)$. If we let $\alpha = d$ and $\beta = -1$, then Theorem 3.2 implies that

$$\det\left(z^d P(z^{-1})\right) = \det\left(z^d A(z^{-1})\right),$$

and we can obtain the infinite eigenvalues of $P(z)$ from $A(z)$.

EXAMPLE 3.3. Let $P(z) = \begin{bmatrix} z^2 & z & i \\ z & z^2 & z \\ -i & z & z \end{bmatrix}$. Applying the Arnoldi process we find

$$V(z) = \begin{bmatrix} 0 & z & \frac{1}{2}z(z^2-z+2i) \\ 1 & 0 & 0 \\ 0 & z & -\frac{1}{2}z(z^2-z+2i) \end{bmatrix}, \; A(z) = \begin{bmatrix} z^2 & 2z^2 & 0 \\ 1 & \frac{1}{2}z(z+1) & \frac{1}{4}(z^4-2z^3+z^2+4) \\ 0 & 1 & \frac{1}{2}z(z+1) \end{bmatrix},$$

such that $P(z)V(z) = V(z)A(z)$. The finite eigenvalues of $P(z)$ are the roots of

$$\det A(z) = z^5 - z^4 - z^3 - z^2.$$

The infinite eigenvalues of $P(z)$ are the zero valued roots of

$$\det\left(z^2 A(z^{-1})\right) = z - z^2 - z^3 - z^4.$$

Since there is one zero valued root of the above polynomial, $P(z)$ has one infinite eigenvalue.

There are several important points from Example 3.3 to make. First, we are able to obtain the finite and infinite eigenvalues of $P(z)$ from $A(z)$. Second, $P(z)$ was self-adjoint, a matrix polynomial is self-adjoint if all of its coefficient matrices are self-adjoint, and the resulting matrix rational was tridiagonal. Finally, the elements of $A(z)$ were all polynomials. It seems that there are starting vectors for the Arnoldi process which force the elements of $A(z)$ to be polynomials. For example, if the starting vector in Example 3.1 were $v = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$, then the elements of $A(z)$ would have been polynomials. Existence and identification of these starting vectors for any matrix polynomial is a topic of future research.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

THEOREM 3.4. *Let $P(z)$ be an $n \times n$ matrix polynomial and $v(z) \in \mathbb{F}^n$ be a nonzero vector such that the vectors*

$$v(z),\, P(z)v(z),\, \ldots,\, P^{n-1}(z)v(z)$$

are linearly independent. Then the Arnoldi process outlined in (3.2) will produce an orthogonal basis, $v_1, v_2, \ldots, v_n$, for $\mathbb{F}^n$. Moreover, the matrix representation of $P(z)$ with respect to this basis is upper Hessenberg, and if $P(z)$ is self-adjoint then the matrix representation is tridiagonal.

Proof. Let $V(z) = \begin{bmatrix} v_1(z) & \cdots & v_n(z) \end{bmatrix}$ and define $A(z) = [a_{ij}(z)]_{i,j=1}^n$, where

$$a_{ij}(z) = \begin{cases} \frac{\langle P(z)v_j(z),\, v_i(z)\rangle}{\langle v_i(z),\, v_i(z)\rangle} & \text{if } j \geq i \\ \beta_{j+1} & \text{if } j = i-1 \\ 0 & \text{if } j < i-1 \end{cases}.$$

Then the columns of $V(z)$ are linearly independent and (3.2) assures that $P(z)V(z) = V(z)A(z)$.

Suppose now that $P(z)$ is self-adjoint, and consider the generalized Fourier coefficients

$$a_{ij}(z) = \frac{\langle P(z)v_j(z),\, v_i(z)\rangle}{\langle v_i(z),\, v_i(z)\rangle},$$

when $j > i$. From Definition 2.1 and the fact that $P(z)$ is self-adjoint we have $\langle P(z)v_j(z),\, v_i(z)\rangle = \langle v_j(z),\, P(z)v_i(z)\rangle$. By (3.2),

$$P(z)v_i(z) = \sum_{k=1}^{i+1} a_{ki}(z)v_k(z).$$

Therefore, $a_{ij}(z) = 0$ whenever $j > i+1$, and it follows that $A(z)$ is tridiagonal. ◻

Let $P(z)$ be an $n \times n$ matrix polynomial and $v(z)$ be a nonzero vector for which Theorem 3.4 holds. Then we can use the Arnoldi process to find an upper Hessenberg matrix $A(z) \in \mathbb{F}^{n\times n}$ that is R-similar to $P(z)$, and by Theorem 3.2 we can obtain the finite and infinite eigenvalues of $P(z)$ from $A(z)$. If the matrix polynomial $P(z)$ is self-adjoint, then (3.2) becomes the three term recurrence.

$$\beta_{j+1}v_{j+1} = \left(P - \frac{\langle Pv_j,\, v_j\rangle}{\langle v_j,\, v_j\rangle} I\right) v_j - \frac{\langle Pv_j,\, v_{j-1}\rangle}{\langle v_{j-1},\, v_{j-1}\rangle} v_{j-1}. \tag{3.3}$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

3.2. Invariant subspaces. We are not guaranteed to be able to form an $n-$dimensional Krylov subspace from every starting vector v. It may happen that k orthogonal vectors have been produced, where $k < n$, and applying (3.2) gives $v_{k+1} = 0$. In this case, the Krylov subspace $K_k(P,\, v)$ is a subspace of $\mathbb{F}^n$ which is invariant under multiplication by P. That is, $PK_k(P, v) \subseteq K_k(P, v)$. We begin with an example of a one dimensional invariant subspace.

EXAMPLE 3.5. Let $P(z) = \begin{bmatrix} z^2+1 & 0 & 1 \\ 0 & z^2+1 & 0 \\ 1 & 0 & z^2+1 \end{bmatrix}$ and $v = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Then

$$P(z)v = (z^2+1)v.$$

It follows that $S = \operatorname{span}\{v\}$ forms a $1-$dimensional subspace of $\mathbb{F}^n$, which is invariant under multiplication by $P(z)$. Moreover, v is an eigenvector corresponding to the eigenvalues $\lambda = \pm i$.

Example 3.5 provides important insight into the eigenvalue problem for matrix polynomials. We have chosen to define finite eigenvalues as the roots of $\det P(z)$, as was done in [1, 2, 4, 6] and the references therein. However, if we consider the matrix polynomial $P(z)$ from Example 3.5 and let $f(z) = z^2+1$, then there is a sense in which $f(z)$ is an eigenvalue of $P(z)$. This idea gains traction when one considers the matrix polynomial $P(z)$ as a linear operator from $\mathbb{F}^n$ to $\mathbb{F}^n$. Moreover, the roots of $f(z)$ are zeros of $\det P(z)$, so there seems to be a correlation between the two competing definitions of an eigenvalue of $P(z)$. Studying these definitions and their relationships is a topic of future research.

For our current discussion, what is important is that invariant subspaces can arise while performing the Arnoldi process, as the following example illustrates.

EXAMPLE 3.6. Let $P(z) = \begin{bmatrix} z^2 & z & 1 & 0 \\ z & z^2 & z & 1 \\ 1 & z & z^2 & z \\ 0 & 1 & z & z^2 \end{bmatrix}$, $v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $v_2 = \begin{bmatrix} -\frac{z}{2} \\ \frac{z}{2} \\ \frac{z}{2} \\ -\frac{z}{2} \end{bmatrix}$.
Then $S = \operatorname{span}\{v_1,\, v_2\}$ is a $2-$dimensional subspace of $\mathbb{F}^n$ which is invariant under multiplication by $P(z)$. In fact, we found this invariant subspace by applying the Arnoldi process to starting vector v_1. Therefore, we have $V(z) = \begin{bmatrix} v_1(z) & v_2(z) \end{bmatrix}$ and

$$A(z) = \begin{bmatrix} 1+\frac{3}{2}z+z^2 & \frac{z^2}{4} \\ 1 & -1-\frac{z}{2}+z^2 \end{bmatrix},$$

such that $P(z)V(z) = V(z)A(z)$.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

Example 3.6 highlights the fact that coming across an invariant subspace while performing the Arnoldi process is a good thing. Since $\det A(z)$ divides $\det P(z)$, see Theorem 3.8, it follows that all the finite eigenvalues of $A(z)$ are also finite eigenvalues of $P(z)$.

3.3. Continue the Arnoldi process. Suppose that after $k < n$ steps of the Arnoldi process outlined in Section 3.1, we find that $v_{k+1} = 0$. Then the Krylov subspace $K_k(P, v)$ is $k-$dimensional subspace of $\mathbb{F}^n$ which is invariant under multiplication by P. We can continue the Arnoldi process, but not with the vector $v_{k+1} = 0$. So, choose v_{k+1} to be a nonzero vector that is orthogonal to each vector in $K_k(P, v)$ and continue on. In this way we can complete our task of finding an orthogonal basis of $\mathbb{F}^n$ such that the matrix representation of P with respect to this basis is upper Hessenberg.

Example 3.7. Here we continue Example 3.6. We choose v_3 to be orthogonal to v_1 and v_2 and the continue the Arnoldi process to obtain

$$V(z) = \begin{bmatrix} 1 & -\frac{z}{2} & \frac{1}{2} & 0 \\ 1 & \frac{z}{2} & 0 & \frac{1}{2}(z-1) \\ 1 & \frac{z}{2} & 0 & -\frac{1}{2}(z-1) \\ 1 & -\frac{z}{2} & -\frac{1}{2} & 0 \end{bmatrix}, \; A(z) = \begin{bmatrix} z^2 + \frac{3z}{2} + 1 & \frac{z^2}{4} & 0 & 0 \\ 1 & z^2 - \frac{z}{2} - 1 & 0 & 0 \\ 0 & 0 & z^2 & (z-1)^2 \\ 0 & 0 & 1 & z(z-1) \end{bmatrix}.$$

Note that in Example 3.7 we effectively split the problem of finding the eigenvalues of $P(z)$ into two smaller problems. This is known as deflation and we summarize this result in the following theorem.

Theorem 3.8. *Let $P(z)$ be a $n \times n$ matrix polynomial. Let $V_1(z) \in \mathbb{F}^{n\times k}$ be a matrix whose columns form a basis for a $k-$dimensional subspace which is invariant under multiplication by $P(z)$. Let $V_2(z) \in \mathbb{F}^{n\times(n-k)}$ be a matrix whose columns are additional vectors, such that the columns of $V(z) = \begin{bmatrix} V_1(z) & V_2(z) \end{bmatrix}$ form a basis of $\mathbb{F}^n$. Let $A(z) \in \mathbb{F}^{n\times n}$ be a matrix such that $P(z)V(z) = V(z)A(z)$, then*

$$A(z) = \begin{bmatrix} A_{11}(z) & A_{12}(z) \\ 0 & A_{22}(z) \end{bmatrix},$$

where $A_{11}(z) \in \mathbb{F}^{k\times k}$ and $A_{22}(z) \in \mathbb{F}^{(n-k)\times(n-k)}$. Moreover, $\det\left(z^\alpha P(z^\beta)\right) = \det\left(z^\alpha A(z^\beta)\right)$ for all $\alpha, \beta \in \mathbb{Z}$.

Proof. The form of $A(z)$ follows readily from the fact that the columns of $V_1(z)$ form a basis for a $k-$dimensional invariant subspace which is invariant under multiplication by $P(z)$. The fact that $\det\left(z^\alpha P(z^\beta)\right) = \det\left(z^\alpha A(z^\beta)\right)$ for all $\alpha, \beta \in \mathbb{Z}$, follows from Theorem 3.2. ◻

Theorem 3.8 can easily be extended to multiple invariant subspaces. Moreover, if the vectors in $V(z)$ come from the Arnoldi process described in Section 3.1, then $A_{11}(z)$ and $A_{22}(z)$ will be upper Hessenberg.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

In conclusion, every $n \times n$ matrix polynomial $P(z)$ is R-similar to an upper Hessenberg $A(z) \in \mathbb{F}^{n\times n}$. Moreover, by Theorem 3.2, the finite eigenvalues of $P(z)$ and $A(z)$ are equal and we can obtain the infinite eigenvalues of $P(z)$ from $A(z)$. Areas of future research include developing efficient numerical methods for computing $A(z)$, and methods for computing subspaces of $\mathbb{F}^n$, which are invariant under multiplication by $P(z)$, directly.

4. Scaling factors and Smith form. In Section 3, we proved that every $n \times n$ matrix polynomial $P(z)$ is R-similar to an upper Hessenberg $A(z)$ in $\mathbb{F}^{n\times n}$. We know that the finite eigenvalues of $P(z)$ are preserved under this transformation. In this section, we assume that $A(z)$ is an upper Hessenberg matrix polynomial. Example 4.1 will show that, in general, the Smith form of $A(z)$ and $P(z)$ are not the same. However, the problems that cause the Smith form to not be preserved seem to be alleviated by a proper choice of scaling factor β_{j+1}, introduced in (3.2). Finally, we provide sufficient conditions under which two similar matrix polynomials have equivalent Smith forms.

EXAMPLE 4.1. Let $P(z) = \begin{bmatrix} z^2 & z & 1 \\ z & z^2 & z \\ 1 & z & z \end{bmatrix}$. Using the Arnoldi process outlined in (3.2), with a scaling factor of 1, we find

$$V(z) = \begin{bmatrix} 0 & z & \frac{1}{2}z^2(z-1) \\ 1 & 0 & 0 \\ 0 & z & -\frac{1}{2}z^2(z-1) \end{bmatrix}, \; A(z) = \begin{bmatrix} z^2 & 2z^2 & 0 \\ 1 & \frac{1}{2}(z^2+z+2) & \frac{1}{4}z^2(z-1)^2 \\ 0 & 1 & \frac{1}{2}(z+2)(z-1) \end{bmatrix},$$

such that $P(z)V(z) = V(z)A(z)$. The Smith Form of $P(z)$ and $A(z)$ are

$$P(z) \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & z-1 & 0 \\ 0 & 0 & z^2(z-1)(z+1) \end{bmatrix}, \; A(z) \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & z^2(z-1)^2(z+1) \end{bmatrix}.$$

Moreover, $P(z)$ has only one infinite eigenvalue, while $A(z)$ has seven infinite eigenvalues.

Example 4.1 shows that, in general, similarity transformations in $\mathbb{F}^{n\times n}$ do not preserve the Smith form of a matrix polynomial. It is important to note that $\det V(z) = 0$ when $z = 0$ and $z = 1$. Moreover, the elementary divisors of $P(z)$ and $A(z)$, corresponding to $z = 1$, are different. We can use our scaling factor β_{j+1} to simplify the columns of $V(z)$, so that this is no longer a problem.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

EXAMPLE 4.2. Let $P(z) = \begin{bmatrix} z^2 & z & 1 \\ z & z^2 & z \\ 1 & z & z \end{bmatrix}$ and $v_1(z) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Using the Arnoldi process, with scaling factor β_{j+1} defined to be the greatest common divisor among the entries of $\hat{v}_{j+1}$ in (3.2), we find

$$V(z) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \; A(z) = \begin{bmatrix} z^2 & 2z & 0 \\ z & \frac{1}{2}(z^2+z+2) & \frac{z}{2}(z-1) \\ 0 & \frac{z}{2}(z-1) & \frac{1}{2}(z+2)(z-1) \end{bmatrix},$$

such that $P(z)V(z) = V(z)A(z)$. Since $\det V(z)$ is a nonzero constant, we know that the Smith form of $A(z)$ and $P(z)$ are the same [4]. In fact, both the finite and infinite elementary divisors of $P(z)$ and $A(z)$ are equal.

Whether or not it is always possible to choose a scaling factor β_{j+1} in (3.2), so that the matrix $A(z)$ has the same Smith form as the original matrix polynomial $P(z)$ is a topic of future research. In what follows we will provide sufficient conditions under which the Smith form of two similar matrix polynomials are equivalent.

Recall that every matrix polynomial $P(z)$ admits the representation

$$P(z) = E(z)D(z)F(z).$$

It is common to refer to the diagonal matrix polynomial $D(z)$ as the Smith form of $P(z)$. The diagonal entries of $D(z)$ are the invariant polynomials. Moreover, if we represent each invariant polynomial as a product of linear factors

$$d_i(z) = (z-\lambda_{i1})^{\alpha_{i1}} \cdots (z-\lambda_{i,k_i})^{\alpha_{i,k_i}},$$

then the factors $(z-\lambda_{ij})^{\alpha_{ij}}$, $j = 1, \ldots, k_i$, $i = 1, \ldots, r$, are called the elementary divisors of $P(z)$ [4]. The invariant polynomials and elementary divisors are closely related to the Jordan chains of a matrix polynomial. In what follows we will use several results from Sections 1.4–1.6 of [4] to prove our main result (Theorem 4.6).

LEMMA 4.3. *Let $P(z)$ be an $n\times n$ matrix polynomial and let $S(z)$ and $V(z)$ be $n\times n$ matrix polynomials such that $S(\lambda)$ and $V(\lambda)$ are nonsingular for some $\lambda \in \mathbb{C}$. Then $y_0, \ldots, y_k$ is a Jordan chain of the matrix polynomial $S(z)P(z)V(z)$ corresponding to λ if and only if the vectors*

$$z_j = \sum_{i=0}^{j} \frac{1}{i!} V^{(i)}(\lambda) y_{j-i}, \quad j = 0, 1, \ldots, k$$

form a Jordan chain of $P(z)$ corresponding to λ.

Proof. See Proposition 1.11 of [4]. ☐

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

From Lemma 4.3, it follows that the matrix polynomials $A(z)$ and $V(z)A(z)$ have the same set of Jordan chains corresponding to $\lambda \in \mathbb{C}$, if $\det V(\lambda) \neq 0$. We will use this fact in the following theorem.

THEOREM 4.4. *Suppose $P(z)V(z) = V(z)A(z)$, where $P(z)$, $V(z)$, $A(z)$ are $n \times n$ matrix polynomials. The length of the Jordan chains of $A(z)$ corresponding to λ are equal to the lengths of the Jordan chains of $P(z)$ corresponding to λ, if $\det V(\lambda) \neq 0$.*

Proof. Let $y_0, \ldots, y_k$ be a Jordan chain of $A(z)$ corresponding to $\lambda \in \mathbb{C}$, where $\det V(\lambda) \neq 0$. By Lemma 4.3, $y_0, \ldots, y_k$ is a Jordan chain of $V(z)A(z)$ corresponding to λ. Therefore, $y_0, \ldots, y_k$ is a Jordan chain of $P(z)V(z)$ corresponding to λ. Again by Lemma 4.3, it follows that the vectors

$$z_j = \sum_{i=0}^{j} \frac{1}{i!} V^{(i)}(\lambda) y_{j-i}, \quad j = 0, 1, \ldots, k$$

form a Jordan chain of $P(z)$ corresponding to λ. ☐

As a consequence of Theorem 4.4 we have the following result.

COROLLARY 4.5. *Let $P(z)V(z) = V(z)A(z)$, where $P(z)$, $V(z)$, $A(z)$ are $n \times n$ matrix polynomials. The degree of the elementary divisors of $A(z)$ corresponding to λ are equal to the degree of the elementary divisors of $P(z)$ corresponding to λ, if $\det V(\lambda) \neq 0$.*

Proof. It follows from Theorem 4.4 that the lengths $k_1, \ldots, k_r$ of the Jordan chains of $A(z)$, corresponding to λ, are equal to the lengths of Jordan chains of $P(z)$, corresponding to λ. Therefore, by Proposition 1.13[1], the nonzero partial multiplicities of $A(z)$ at λ are equal to the nonzero partial multiplicities of $P(z)$ at λ. Since the nonzero partial multiplicities coincide with the degrees of the elementary divisors corresponding to λ [4], the result follows. ☐

One can use the number of invariant polynomials and the elementary divisors to uniquely determine the Smith form of a matrix polynomial. We will use this fact to prove the following.

THEOREM 4.6. *Let $P(z)$ be a regular $n \times n$ matrix polynomial. Let $A(z)$ be a matrix polynomial that is R-similar to $P(z)$. Then the Smith form of $P(z)$ and $A(z)$ are equivalent, if none of the finite eigenvalues of $P(z)$ are also eigenvalues of $V(z)$.*

Proof. Since $P(z)$ is regular, we know that $\det P(z)$ is nonzero and therefore $P(z)$ has n invariant polynomials. The matrix polynomial $A(z)$ is R-similar to $P(z)$, therefore $A(z)$ also has n invariant polynomials. Let $\lambda \in \mathbb{C}$ be an eigenvalue of $A(z)$,

[1] Many results in Sections 1.4–1.6 of [4], including Proposition 1.13, have been stated for monic matrix polynomials, but monotonicity is not used in their proofs.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 31, pp. 321-334, May 2016

ELA
http:/repository.uwyo.edu/ela

then λ is also an eigenvalue of $P(z)$. If the $\det V(\lambda)$ is nonzero, then it follows from Corollary 4.5 that the elementary divisors of $A(z)$ and $P(z)$ corresponding to λ are the same. Therefore, if none of the finite eigenvalues of $P(z)$ are also eigenvalues of $V(z)$, then $A(z)$ and $P(z)$ have the same number of invariant polynomials and the same elementary divisors. It follows that the Smith form of $P(z)$ and $A(z)$ are equivalent. $\square$

In conclusion, we have introduced an pseudo inner product suitable for the vector space $\mathbb{F}^n$, where $\mathbb{F}$ is the field of rational functions. Armed with this pseudo inner product, we defined orthogonality of vectors in $\mathbb{F}^n$ and showed that we can apply the Arnoldi method directly to matrix polynomials. This allowed us to construct an upper Hessenberg matrix in $\mathbb{F}^{n\times n}$ that is similar to the original matrix polynomial. We proved that this transformation preserves the determinant, and we saw scenarios in which the finite and infinite elementary divisors are also preserved. Future research includes developing numerical methods for computing this transformation efficiently, finding optimal starting vectors for the Arnoldi process, and determining if we can always do this transformation in such a way that preserves the Smith form of the original matrix polynomial.

Acknowledgments. The author wishes to thank David Watkins for several conversations which helped construct the ideas in this paper, and also thank Peter Lancaster for suggestions which led to improvements.

REFERENCES

[1] D.A. Bini and V. Noferini. Solving polynomial eigenvalue problem by means of the Ehrlich-Aberth method. *Linear Algebra Appl.*, 439:1130–1149, 2013.
[2] D.A. Bini, V. Noferini, and M. Sharify. Locating the eigenvalues of matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 34:1708–1727, 2013.
[3] J. Gary. Hyman's method applied to the general eigenvalue problem. *Math. Comp.*, 19:314–316, 1965.
[4] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials.* SIAM, Philadelphia, PA, 2009.
[5] L. Hoffnung, R.C. Li, and Q. Ye. Krylov type subspace methods for matrix polynomials. *Linear Algebra Appl.*, 415:52–81, 2006.
[6] B. Plestenjak. Numerical methods for the tridiagonal hyperbolic quadratic eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 28:1157–1172, 2006.
[7] L. Taslaman, F. Tisseur, and I. Zaballa. Triangularizing matrix polynomials. *Linear Algebra Appl.*, 439:1679–1699, 2013.
[8] F. Tisseur and I. Zaballa. Triangularizing quadratic matrix polynomials. *SIAM J. Matrix Anal. Appl.*, 34:312–337, 2013.
[9] D.S. Watkins. Some perspectives on the eigenvalue problem. *SIAM Rev.*, 35:430–471, 1993.
[10] J.H. Wilkinson. *Rounding Errors in Algebraic Processes.* Prenctice-Hall, New Jersey, 1963.