Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

A FACTORIZATION OF THE INVERSE OF THE SHIFTED COMPANION MATRIX*

JARED L. AURENTZ†

Abstract. A common method for computing the zeros of an n-th degree polynomial is to compute the eigenvalues of its companion matrix C. A method for factoring the shifted companion matrix $C - \rho I$ is presented. The factorization is a product of $2n - 1$ essentially 2×2 matrices and requires $O(n)$ storage. Once the factorization is computed it immediately yields factorizations of both $(C - \rho I)^{-1}$ and $(C - \rho I)^*$. The cost of multiplying a vector by $(C - \rho I)^{-1}$ is shown to be $O(n)$, and therefore, any shift and invert eigenvalue method can be implemented efficiently. This factorization is generalized to arbitrary forms of the companion matrix.

Key words. Polynomial, Root, Companion matrix, Shift and invert.

AMS subject classifications. 65F18, 15A18.

1. Introduction. Finding the roots of an n-th degree polynomial is equivalent to finding the eigenvalues of the corresponding companion matrix. It has been shown in [1] that the companion matrix can be factored into a product of essentially 2×2 matrices. Moreover, these factors can be reordered arbitrarily to produce different forms of the companion matrix. Recently, methods have been proposed in [2] that exploit this structure to compute the entire set of roots in $O(n^2)$ time.

For large n computing the entire set of roots may be infeasible. One option for computing a subset of the roots is to use a shift and invert strategy. We will show that the shifted companion matrix, the inverse of the shifted companion matrix and the adjoints of both these matrices can be written in a factored form that requires $O(n)$ storage and can be applied to a vector in $O(n)$ operations. This means that the inverse power method or any subspace iteration method can be used to compute a modest subset of the spectrum in $O(n)$ time.

In Section 2, we introduce the idea of core transformation and describe the Fielder factorization. In Section 3, we state and prove the main factorization theorems. In Section 4, we extend the factorization to related operators. In Section 5, we give some numerical examples that illustrate the use of the factorization to compute a few roots of a polynomial of high degree.

*Received by the editors on August 23, 2012. Accepted for publication on May 4, 2013. Handling Editor: Bryan L. Shader.

†Department of Mathematics, Washington State University, Pullman, WA 99164-3113, USA (jaurentz@math.wsu.edu).

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

2. The factored companion matrix. In order to describe the Fiedler factorization we introduce the idea of a core transformation. A *core transformation* is any matrix G_i that is the identity matrix everywhere except at the intersection of rows and columns i and $i+1$. These matrices are essentially 2×2 and can be viewed as transformations that act on either two adjacent rows or two adjacent columns. Graphically we represent a core transformation with an arrowed bracket,

$$\begin{array}{c} \text{↱} \\ \text{↳} \\ \\ \\ \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix},$$

where the arrowheads represent the rows that are transformed. It is easy to see that two core transformations G_i and G_j commute, $G_iG_j = G_jG_i$ as long as $|i-j| > 1$. This can be seen graphically as two brackets that don't overlap,

$$\begin{array}{cc} \text{↱} & \\ \text{↳} & \\ & \text{↱} \\ & \text{↳} \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} = \begin{array}{cc} & \text{↱} \\ & \text{↳} \\ \text{↱} & \\ \text{↳} & \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} = \begin{array}{c} \text{↱} \\ \text{↳} \\ \text{↱} \\ \text{↳} \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}.$$

If $|i-j| \leq 1$, then the factors do not commute in general. This appears graphically as overlapping brackets,

$$\begin{array}{cc} \text{↱} & \\ \text{↳} & \text{↱} \\ & \text{↳} \\ & \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \neq \begin{array}{cc} & \text{↱} \\ \text{↱} & \text{↳} \\ \text{↳} & \\ & \end{array} \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}.$$

Fiedler showed that for an n-th degree, monic polynomial

$$p(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$$

the associated upper-Hessenberg companion matrix

$$C = \begin{bmatrix} -a_1 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}$$

can be written as a product of n core transformations, $C = C_1C_2\cdots C_{n-1}C_n$. For $1 \leq i \leq n-1$, the C_i have the form

$$C_i = \begin{bmatrix} I_{i-1} & & & \\ & -a_i & 1 & \\ & 1 & 0 & \\ & & & I_{n-i-1} \end{bmatrix}, \tag{2.1}$$

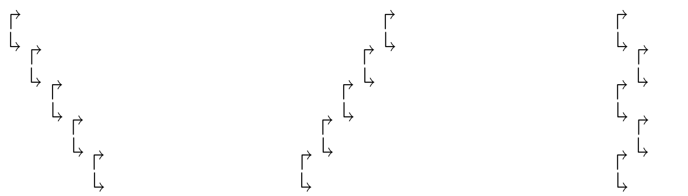
Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

where I_k is the $k \times k$ identity matrix. For $i = n$, there is a special factor, $C_n = \operatorname{diag}\{1, \ldots, 1, -a_n\}$. Fiedler also showed that if the factors $\{C_1, C_2, \ldots, C_n\}$ are multiplied together in an arbitrary order, the resulting product is similar to C and therefore can also be considered a companion matrix for $p(x)$. In any such product, many of the factors commute, so we only need to keep track of the ones that don't. To do this we introduce the *position vector* $\mathbf{p}$. This is an $(n-1)$-tuple that takes the value $p_i = 0$ if C_i is to the left of C_{i+1}, and $p_i = 1$ if C_{i+1} is to the left of C_i.

The products of the C_i are not unique to the *position vector* $\mathbf{p}$ since C_i and C_j commute if $|i - j| > 1$. This means the *position vector* $\mathbf{p}$ represents an equivalence class of companion matrices. For example, when $n = 4$, the product $C = C_1C_3C_2C_4$ is equivalent to the products $C_3C_1C_2C_4$, $C_1C_3C_4C_2$ and $C_3C_1C_4C_2$ since they are all consistent with the same *position vector* $\mathbf{p} = \{0, 1, 0\}$.

Definition 2.1. Let $\{G_1, G_2, \ldots, G_n\}$ be a set of core transformations and let $\mathbf{p}$ be a *position vector* that describes the relative order of the G_i. Any product of these core transformations that is consistent with $\mathbf{p}$ will be written as

$$\prod_{\mathbf{p}} \{G_1, G_2, \ldots, G_n\}.$$



(a) Upper-Hessenberg (b) Lower-Hessenberg (c) CMV-pattern

Fig. 2.1: Graphical representation of some factorizations.

Example 2.2. (Upper-Hessenberg matrix). For the upper-Hessenberg companion matrix the factorization has position vector $\mathbf{p} = \{0, 0, \ldots, 0\}$. This means that C can be written as $C = \prod_{\mathbf{p}}\{C_1, C_2, \ldots, C_n\} = C_1C_2 \cdots C_n$. This is represented graphically in Figure 2.1a as a descending sequence of core transformations.

Example 2.3. (Lower-Hessenberg matrix). For the lower-Hessenberg companion matrix the factorization has position vector $\mathbf{p} = \{1, 1, \ldots, 1\}$. This means that C can be written as $C = \prod_{\mathbf{p}}\{C_1, C_2, \ldots, C_n\} = C_nC_{n-1} \cdots C_1$. This is represented graphically in Figure 2.1b as an ascending sequence of core transformations.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

Example 2.4. (CMV matrix). For the CMV companion matrix the factorization has position vector $\mathbf{p} = \{0, 1, 0, 1, \ldots\}$. This means that C can be written as $C = \prod_{\mathbf{p}}\{C_1, C_2, \ldots, C_n\} = (C_1C_3\cdots)(C_2C_4\cdots)$. This is represented graphically in Figure 2.1c as a zigzag sequence of core transformations.

3. Factorization theorems. In this section, we will show that for the upper-Hessenberg companion matrix C there exists a sequence of $2n-1$ core transformations whose product equals $C - \rho I$. We will then extend the result to companion matrices with arbitrary *position vector* $\mathbf{p}$. Before we state and prove the main results we introduce the *Horner vector*, $\mathbf{h}(\rho) \in \mathbb{C}^n$. This vector contains all the intermediate steps used when applying *Horner's rule* to the polynomial p, at the shift ρ. The entries are defined recursively,

$$h_1(\rho) = \rho + a_1, \quad h_{i+1}(\rho) = \rho h_i(\rho) + a_{i+1}. \tag{3.1}$$

Notice that $h_n(\rho) = p(\rho)$. We also introduce two new types of core transformations:

$$R_i = \begin{bmatrix} I_{i-1} & & & \\ & 1 & -\rho & \\ & 0 & 1 & \\ & & & I_{n-i-1} \end{bmatrix} \tag{3.2}$$

for $i = 1, \ldots, n-1$,

$$H_i = \begin{bmatrix} I_{i-1} & & & \\ & -h_i(\rho) & 1 & \\ & 1 & 0 & \\ & & & I_{n-i-1} \end{bmatrix} \tag{3.3}$$

for $i = 1, \ldots, n-1$, and $H_n = \operatorname{diag}\{I_{n-1}, -h_n(\rho)\}$. Here I_k is still the $k \times k$ identity matrix.

Theorem 3.1. *For an n-th degree, monic polynomial, $p(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$, with $a_n \neq 0$, the shifted, upper-Hessenberg companion matrix C associated with $p(x)$, can be written as the product of $2n-1$ core transformations,*

$$C - \rho I = H_1H_2\cdots H_{n-2}H_{n-1}H_nR_{n-1}R_{n-2}\cdots R_2R_1,$$

with H_i and R_i defined as in (3.2) and (3.3), respectively.

Proof. The case $n = 1$ is trivial, $C - \rho I = [-a_1 - \rho] = [-h_1(\rho)] = H_n$. For $n > 1$,

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

assume that after $k-1$ steps, we have

$$C - \rho I = H_1 \dots H_{k-1} \begin{bmatrix} I_{k-1} & & & \\ & -h_k(\rho) & -a_{k+1} & \cdots \\ & 1 & -\rho & \\ & & \ddots & \ddots \end{bmatrix} R_{k-1} \cdots R_1.$$

We compute the next factor by first performing an interchange of rows k and $k+1$,

$$\begin{bmatrix} I_{k-1} & & & \\ & -h_k(\rho) & -a_{k+1} & \cdots \\ & 1 & -\rho & \\ & & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} I_{k-1} & & & \\ & 0 & 1 & \\ & 1 & 0 & \\ & & & I_{n-k-1} \end{bmatrix} \begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & -h_k(\rho) & -a_{k+1} & \cdots \\ & & \ddots & \ddots \end{bmatrix}.$$

Then we use the correct Gauss transformation to eliminate the $(k+1, k)$ entry,

$$\begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & -h_k(\rho) & -a_{k+1} & \cdots \\ & & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} I_{k-1} & & & \\ & 1 & 0 & \\ & -h_k(\rho) & 1 & \\ & & & I_{n-k-1} \end{bmatrix} \begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & 0 & -h_{k+1}(\rho) & \cdots \\ & & \ddots & \ddots \end{bmatrix}.$$

Combining the row interchange with the Gauss transform we get the factor H_k,

$$\begin{bmatrix} I_{k-1} & & & \\ & -h_k(\rho) & -a_{k+1} & \cdots \\ & 1 & -\rho & \\ & & \ddots & \ddots \end{bmatrix} = H_k \begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & 0 & -h_{k+1}(\rho) & \cdots \\ & & \ddots & \ddots \end{bmatrix}.$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

The last step is to eliminate the $(k, k+1)$ entry using another Gauss transform,

$$\begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & 0 & -h_{k+1}(\rho) & \cdots \\ & & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} I_{k-1} & & & \\ & 1 & 0 & \\ & 0 & -h_{k+1}(\rho) & \cdots \\ & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & 0 & 1 & \\ & & & I_{n-k-1} \end{bmatrix}.$$

This core transformation is exactly R_k,

$$\begin{bmatrix} I_{k-1} & & & \\ & 1 & -\rho & \\ & 0 & -h_{k+1}(\rho) & \cdots \\ & & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} I_{k-1} & & & \\ & 1 & 0 & \\ & 0 & -h_{k+1}(\rho) & \cdots \\ & & \ddots & \ddots \end{bmatrix} R_k.$$

Putting it all together we get

$$\begin{bmatrix} I_{k-1} & & & \\ & -h_k(\rho) & -a_{k+1} & \cdots \\ & 1 & -\rho & \\ & & \ddots & \ddots \end{bmatrix} = H_k \begin{bmatrix} I_k & & & \\ & -h_{k+1}(\rho) & -a_{k+2} & \cdots \\ & 1 & -\rho & \\ & & \ddots & \ddots \end{bmatrix} R_k,$$

which implies that

$$C - \rho I = H_1 \cdots H_k \begin{bmatrix} I_k & & & \\ & -h_{k+1}(\rho) & -a_{k+2} & \cdots \\ & 1 & -\rho & \\ & & \ddots & \ddots \end{bmatrix} R_k \cdots R_1. \quad \square$$

EXAMPLE 3.2. (Upper-Hessenberg factors). When $n = 3$ the shifted, upper-Hessenberg companion matrix admits the factorization $C - \rho I = H_1 H_2 H_3 R_2 R_1$,

$$\begin{bmatrix} -h_1(\rho) & -a_2 & -a_3 \\ 1 & -\rho & \\ & 1 & -\rho \end{bmatrix} =$$

$$\begin{bmatrix} -h_1(\rho) & 1 & \\ 1 & 0 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & -h_2(\rho) & 1 \\ & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & -h_3(\rho) \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & -\rho \\ & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\rho & \\ 0 & 1 & \\ & & 1 \end{bmatrix}.$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

It should be noted that this factorization is a natural extension of the Fiedler factorization in the sense that when $\rho = 0$, $H_i = C_i$ and $R_i = I$ for all i,

$$C - 0I = H_1 \cdots H_{n-1} H_n R_{n-1} \cdots R_1 = C_1 \cdots C_{n-1} C_n.$$

In order to extend the factorization to companion matrices with arbitrary *position vectors* $\mathbf{p}$, we need the following lemma.

LEMMA 3.3. *For an n-th degree, monic polynomial, $p(x) = x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n$, with $a_n \neq 0$, the companion matrix C associated with $p(x)$ and position vector $\mathbf{p}$, has only one non-zero diagonal entry, namely $C_{1,1} = -a_1$.*

Proof. The case $n = 1$ and $n = 2$ are trivial. For the induction step it is sufficient to note that the k-th diagonal entry of C is uniquely determined by the factors C_{k-1} and C_k and the $(k-1)$ entry of $\mathbf{p}$. This is true because core transformations only act locally. For $p_{k-1} = 0$, the k-th diagonal of C is the k-th diagonal entry of $C_{k-1}C_k$,

$$C_{k-1}C_k = \begin{bmatrix} I_{k-2} & & & & \\ & -a_{k-1} & -a_k & 1 & \\ & 1 & \mathbf{0} & 0 & \\ & 0 & 1 & 0 & \\ & & & & I_{n-k-1} \end{bmatrix}.$$

The k-th diagonal entry is $\mathbf{0}$. For $p_{k-1} = 1$, the k-th diagonal of C is the k-th diagonal entry of $C_k C_{k-1}$,

$$C_k C_{k-1} = \begin{bmatrix} I_{k-2} & & & & \\ & -a_{k-1} & 1 & 0 & \\ & -a_k & \mathbf{0} & 1 & \\ & 1 & 0 & 0 & \\ & & & & I_{n-k-1} \end{bmatrix}. \qquad \square$$

The importance of Lemma 3.3 is that every companion matrix formed via the Fiedler factors has the same main diagonal.

DEFINITION 3.4. Let $P_i \in \mathbb{C}^{n \times n}$ for $i = 1, \ldots, n$ be the following diagonal matrix

$$P_i = \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & I_{n-i} \end{bmatrix}.$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

When $i = 0$, $P_i = I_n$.

Corollary 3.5. *For an n-th degree, monic polynomial, $p(x) = x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n$, with $a_n \neq 0$, the shifted, companion matrix C associated with $p(x)$ and position vector $\mathbf{p}$, can be written as*

$$\prod_{\mathbf{p}}\{H_1, C_2, \ldots, C_n\} - \rho P_1.$$

Proof. By Lemma 3.3 we know the main diagonal of $C - \rho I$,

$$C - \rho I = \begin{bmatrix} -a_1 - \rho & * & \cdots & * \\ * & -\rho & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \cdots & * & -\rho \end{bmatrix} = \begin{bmatrix} -h_1(\rho) & * & \cdots & * \\ * & -\rho & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \cdots & * & -\rho \end{bmatrix}.$$

Now we subtract off ρP_1 and leave the rest of $C - \rho I$ in the matrix D,

$$C - \rho I = D - \rho P_1$$

where

$$D = \begin{bmatrix} -h_1(\rho) & * & \cdots & * \\ * & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \cdots & * & 0 \end{bmatrix}.$$

Since C and D only differ in the $(1,1)$ entry D has the same off diagonal entries as C. The $(1,1)$ entry of C was determined by the factor C_1. Replacing the $(1,1)$ entry of C_1 with $-h_1(\rho)$ transforms C_1 into H_1. Therefore,the matrix D can be written as

$$D = \prod_{\mathbf{p}}\{H_1, C_2, \ldots, C_n\},$$

and therefore,

$$C - \rho I = \prod_{\mathbf{p}}\{H_1, C_2, \ldots, C_n\} - \rho P_1. \qquad \square$$

Theorem 3.6. *For an n-th degree, monic polynomial, $p(x) = x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n$, with $a_n \neq 0$, the shifted, companion matrix C associated with $p(x)$ and position vector $\mathbf{p}$, can be written as the product of $2n-1$ core transformations,*

$$C - \rho I = A_1 A_2 \cdots A_{n-2} A_{n-1} H_n B_{n-1} B_{n-2} \cdots B_2 B_1,$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

where $A_i = H_i$ and $B_i = R_i$ when $p_i = 0$ and $A_i = R_i^T$ and $B_i = H_i$ when $p_i = 1$ with H_i and R_i defined as in (3.2) and (3.3), respectively.

Proof. The case $n = 1$ is the same as in Theorem 3.1. For $n > 1$, the induction step has two cases depending on the value of p_k. For the first case, we start by assuming that $p_k = 0$ and after $k-1$ steps $C - \rho I$ has been factored as

$$C - \rho I = A_1 \cdots A_{k-1} \left(\prod_{\mathbf{p}} \{H_k, C_{k+1}, \ldots, C_n\} - \rho P_k \right) B_{k-1} \cdots B_1,$$

where the A_i and B_i are chosen exactly as in the statement of the theorem. We begin by removing the factor H_k exactly as in Theorem 3.1. This does not affect the rest of the C_i for $i > k+1$ because non-intersecting core transformations commute,

$$A_1 \cdots A_{k-1} H_k \left(\prod_{\mathbf{p}} \{I, C_{k+1}, \ldots, C_n\} - \rho H_k^{-1} P_k \right) B_{k-1} \cdots B_1.$$

Let's take a closer look at the product $\rho H_k^{-1} P_k$,

$$\rho H_k^{-1} P_k = \begin{bmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & 0 & \rho & & \\ & & 0 & \rho h_k(\rho) & & \\ & & & & \rho & \\ & & & & & \ddots \end{bmatrix}.$$

This product can be written as

$$\rho H_k^{-1} P_k = F_k + \rho P_{k+1},$$

where

$$F_k = \begin{bmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & 0 & \rho & & \\ & & 0 & \rho h_k(\rho) & & \\ & & & & 0 & \\ & & & & & \ddots \end{bmatrix}.$$

Since F_k is non-zero only at the intersection of rows and columns k and $k+1$, the only C_i in the product $\prod_{\mathbf{p}} \{I, C_{k+1}, \ldots, C_n\}$ that F_k interacts with is C_{k+1}. Computing

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

the sum we get

$$C_{k+1} - F_k = \begin{bmatrix} I_{k-1} & & & & \\ & 1 & -\rho & 0 & \\ & 0 & -h_{k+1}(\rho) & 1 & \\ & 0 & 1 & 0 & \\ & & & & I_{n-k-2} \end{bmatrix}.$$

Multiplying by the correct Gauss transform on the right to eliminate the $(k, k+1)$ entry we get

$$C_{k+1} - F_k = \begin{bmatrix} I_{k-1} & & & & \\ & 1 & 0 & 0 & \\ & 0 & -h_{k+1}(\rho) & 1 & \\ & 0 & 1 & 0 & \\ & & & & I_{n-k-2} \end{bmatrix} \begin{bmatrix} I_{k-1} & & & & \\ & 1 & -\rho & 0 & \\ & 0 & 1 & 0 & \\ & 0 & 0 & 1 & \\ & & & & I_{n-k-2} \end{bmatrix}$$

which is exactly

$$C_{k+1} - F_k = H_{k+1} R_k.$$

Putting this all together we obtain

$$H_k \left(\textstyle\prod_{\mathbf{p}} \{I, C_{k+1}, \ldots, C_n\} - \rho H_k^{-1} P_k \right) = \\ H_k \left(\textstyle\prod_{\mathbf{p}} \{H_{k+1}, C_{k+2}, \ldots, C_n\} R_k - \rho P_{k+1} \right),$$

and since $P_{k+1} R_k^{-1} = P_{k+1}$, we obtain

$$H_k \left(\textstyle\prod_{\mathbf{p}} \{I, C_{k+1}, \ldots, C_n\} - \rho H_k^{-1} P_k \right) = \\ H_k \left(\textstyle\prod_{\mathbf{p}} \{H_{k+1}, C_{k+2}, \ldots, C_n\} - \rho P_{k+1} \right) R_k.$$

Using the induction hypothesis we get our result,

$$C - \rho I = \\ A_1 \cdots A_{k-1} H_k \left(\textstyle\prod_{\mathbf{p}} \{H_{k+1}, C_{k+2}, \ldots, C_n\} - \rho P_{k+1} \right) R_k B_{k-1} \cdots B_1.$$

For the second case we start by assuming that $p_k = 1$ and after $k-1$ steps $C - \rho I$ has been factored as

$$C - \rho I = A_1 \cdots A_{k-1} \left(\prod_{\mathbf{p}} \{H_k, C_{k+1}, \ldots, C_n\} - \rho P_k \right) B_{k-1} \cdots B_1.$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

We begin by removing the factor H_k but this time we factor on the right since $p_k = 1$ tells us that H_k is to the right of C_{k+1},

$$A_1 \cdots A_{k-1} \left(\prod_{\mathbf{p}} \{I, C_{k+1}, \ldots, C_n\} - \rho P_k H_k^{-1} \right) H_k B_{k-1} \cdots B_1.$$

Again taking a closer look at the product $\rho P_k H_k^{-1}$,

$$\rho P_k H_k^{-1} = \begin{bmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & 0 & 0 & & \\ & & \rho & \rho h_k(\rho) & & \\ & & & & \rho & \\ & & & & & \ddots \end{bmatrix}.$$

This product can be written as

$$\rho P_k H_k^{-1} = F_k^T + \rho P_{k+1}$$

where

$$F_k^T = \begin{bmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & 0 & 0 & & \\ & & \rho & \rho h_k(\rho) & & \\ & & & & 0 & \\ & & & & & \ddots \end{bmatrix}.$$

Since F_k^T is non-zero only at the intersection of rows and columns k and $k+1$, the only C_i in the product $\prod_{\mathbf{p}} \{I, C_{k+1}, \ldots, C_n\}$ that F_k interacts with is C_{k+1}. Computing the sum we get

$$C_{k+1} - F_k^T = \begin{bmatrix} I_{k-1} & & & & \\ & 1 & 0 & 0 & \\ & -\rho & -h_{k+1}(\rho) & 1 & \\ & 0 & 1 & 0 & \\ & & & & I_{n-k-2} \end{bmatrix}.$$

Multiplying by the correct Gauss transform on the left to eliminate the $(k+1, k)$ entry we get

$$C_{k+1} - F_k^T = \begin{bmatrix} I_{k-1} & & & & \\ & 1 & 0 & 0 & \\ & -\rho & 1 & 0 & \\ & 0 & 0 & 1 & \\ & & & & I_{n-k-2} \end{bmatrix} \begin{bmatrix} I_{k-1} & & & & \\ & 1 & 0 & 0 & \\ & 0 & -h_{k+1}(\rho) & 1 & \\ & 0 & 1 & 0 & \\ & & & & I_{n-k-2} \end{bmatrix}$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

which is exactly

$$C_{k+1} - F_k^T = R_k^T H_{k+1}.$$

Putting this all together we get

$$\left(\prod\nolimits_{\mathbf{p}}\{I, C_{k+1}, \ldots, C_n\} - \rho P_k H_k^{-1}\right) H_k = \left(R_k^T \prod\nolimits_{\mathbf{p}}\{H_{k+1}, C_{k+2}, \ldots, C_n\} - \rho P_{k+1}\right) H_k,$$

and since $(R_k^T)^{-1} P_{k+1} = P_{k+1}$, we obtain

$$\left(\prod\nolimits_{\mathbf{p}}\{I, C_{k+1}, \ldots, C_n\} - \rho P_k H_k^{-1}\right) H_k = R_k^T \left(\prod\nolimits_{\mathbf{p}}\{H_{k+1}, C_{k+2}, \ldots, C_n\} - \rho P_{k+1}\right) H_k.$$

Using the induction hypothesis we get our result,

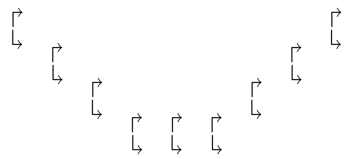
$$C - \rho I = A_1 \cdots A_{k-1} R_k^T \left(\prod\nolimits_{\mathbf{p}}\{H_{k+1}, C_{k+2}, \ldots, C_n\} - \rho P_{k+1}\right) H_k B_{k-1} \cdots B_1. \quad \square$$

It should be noted that Theorem 3.1 is a special case of Theorem 3.6.

Example 3.7. (CMV factors). When $n = 3$ the shifted, CMV companion matrix with $\mathbf{p} = \{1, 0\}$ admits the factorization $C - \rho I = R_1^T H_2 H_3 R_2 H_1$,

$$\begin{bmatrix} -h_1(\rho) & 1 & 0 \\ -a_2 & -\rho & -a_3 \\ 1 & & -\rho \end{bmatrix} = \begin{bmatrix} 1 & 0 & \\ -\rho & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & -h_2(\rho) & 1 \\ & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & -h_3(\rho) \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & -\rho \\ & 0 & 1 \end{bmatrix} \begin{bmatrix} -h_1(\rho) & 1 & \\ 1 & 0 & \\ & & 1 \end{bmatrix}.$$

Schematically, the factorization looks like

$(C - \rho I) =$ 

when $n = 5$. The only difference from one shifted companion matrix to another is which side of the *"V"* a particular H_i, R_i or R_i^T resides.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

4. Extensions. In this section, we will see that this factorization gives easy access to some other operators. Remember what $C - \rho I$ looks like,

$$C - \rho I = A_1 \cdots A_{n-1} H_n B_{n-1} \cdots B_1.$$

In order to use a shift and invert strategy to compute the roots it is necessary to have access to the inverse of the shifted operator. Here we get the factored inverse of the shifted matrix for free,

$$(C - \rho I)^{-1} = B_1^{-1} \cdots B_{n-1}^{-1} H_n^{-1} A_{n-1}^{-1} \cdots A_1^{-1}.$$

EXAMPLE 4.1. (CMV factors). When $n = 3$ the inverse of the shifted, CMV companion matrix with $\mathbf{p} = \{1, 0\}$ admits the factorization

$$(C - \rho I)^{-1} = (H_1)^{-1}(R_2)^{-1}(H_3)^{-1}(H_2)^{-1}(R_1^T)^{-1},$$

$$h_3^{-1}(\rho) \begin{bmatrix} -\rho^2 & -\rho & a_3 \\ \rho a_2 + a_3 & -\rho h_1(\rho) & h_1(\rho) a_3 \\ -\rho & -1 & -h_2(\rho) \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 1 & \\ 1 & h_1(\rho) & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \rho \\ & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & -h_3^{-1}(\rho) \end{bmatrix} \begin{bmatrix} 1 & & \\ & 0 & 1 \\ & 1 & h_2(\rho) \end{bmatrix} \begin{bmatrix} 1 & 0 & \\ \rho & 1 & \\ & & 1 \end{bmatrix}.$$

Some sparse eigenvalue methods, for instance unsymmetric-Lanczos, require access to the adjoint. We also get the factored adjoint for free,

$$(C - \rho I)^* = B_1^* \cdots B_{n-1}^* H_n^* A_{n-1}^* \cdots A_1^*.$$

The adjoint of the inverse is available as well,

$$((C - \rho I)^{-1})^* = (A_1^{-1})^* \cdots (A_{n-1}^{-1})^* (H_n^{-1})^* (B_{n-1}^{-1})^* \cdots (B_1^{-1})^*.$$

It is interesting to note that all of these operators have a similar sparsity structure. More over the cost of applying any of these operators to a vector is the same. The cost to compute the Horner vector is roughly $2n$. Once the Horner vector is computed each of the H_i and R_i or their inverses, transposes or adjoints must be applied to the vector. Since each of these core transformations requires 2 flops, except H_n, the flop count sums to roughly $4n$. So the total flop count for applying the operator is roughly $6n$ and only $4n$ if the Horner vector has already been computed.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

5. Numerical experiments. In this section, we give two examples of using a shift and invert strategy to compute subsets of the roots of a polynomial of high degree.

To compute the roots we used the shifted inverse of the upper-Hessenberg companion matrix in conjunction with a version of the unsymmetric-Lanczos algorithm as outlined in [3]. The scaling coefficients were chosen such that the resulting tridiagonal matrix was complex symmetric. Its eigenvalues could then be computed using the complex symmetric eigensolver described in [4]. Implicit restarts were performed using a Krylov-Schur-like method described in [3].

For the first example, we computed the 10 roots of the polynomial $p(x) = x^{10000} - i$ that are closest to 1. Since these roots are known, we can compute the relative error exactly.

λ	$\kappa(\lambda)$	$\frac{\lvert\lambda-\mu\rvert}{\lvert\mu\rvert}$
λ_1	1.0×10^0	7.0×10^{-19}
λ_2	1.0×10^0	1.3×10^{-16}
λ_3	1.0×10^0	2.0×10^{-18}
λ_4	1.0×10^0	2.7×10^{-16}
λ_5	1.0×10^0	3.9×10^{-18}
λ_6	1.0×10^0	5.1×10^{-16}
λ_7	1.0×10^0	4.5×10^{-17}
λ_8	1.0×10^0	6.4×10^{-16}
λ_9	1.0×10^0	3.7×10^{-16}
λ_{10}	1.0×10^0	6.6×10^{-16}

Table 5.1: Error table for computed roots of $x^{10000} - i$.

Table 5.1 shows the relative errors of the computed root λ against the exact root μ. Since we are running unsymmetric-Lanczos, we can compute a condition number for the roots. If v is the right eigenvector and w is the left eigenvector associated with λ then $\kappa(\lambda)$ is defined as

$$\kappa(\lambda) = \frac{||v||_2||w||_2}{|w^*v|}.$$

Figure 5.1 shows the 10 computed roots of $i = \sqrt{-1}$ closest to 1 against the 20 actual roots of $i = \sqrt{-1}$ closest to 1. In this example, the shift and invert strategy computed the 10 closest roots without leaving any gaps. For the second example, we

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

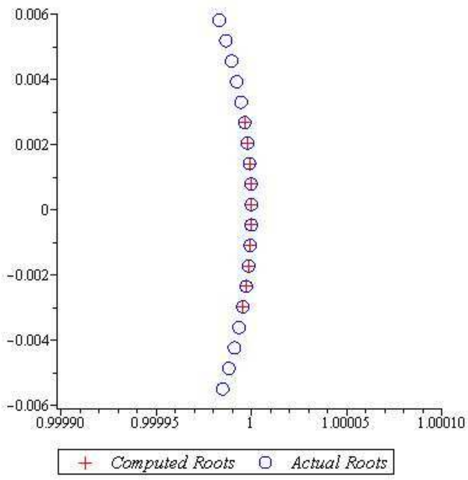


Fig. 5.1: Computed Roots of $x^{10000} - i$ against Actual Roots.

compute the 10 roots closest to $i = \sqrt{-1}$ of a millionth degree, complex polynomial with coefficients whose real and imaginary parts are both normally distributed with mean 0 and variance 1.

λ_1	$-2.33602533173235511 \times 10^{-6}$	$+$	$i1.0000010175657970$
λ_2	$7.68463687578200042 \times 10^{-6}$	$+$	$i0.99999830744672802$
λ_3	$1.64702381595330610 \times 10^{-6}$	$+$	$i0.99998902848130167$
λ_4	$-1.29532226619050469 \times 10^{-5}$	$+$	$i1.0000011072894426$
λ_5	$1.30334397644394113 \times 10^{-5}$	$+$	$i1.0000000425872966$
λ_6	$-2.05670539367284743 \times 10^{-5}$	$+$	$i1.0000049090149605$
λ_7	$-1.98408381080531119 \times 10^{-5}$	$+$	$i0.99998890776490701$
λ_8	$2.32231047287246357 \times 10^{-5}$	$+$	$i1.0000002882809333$
λ_9	$3.05183754248373172 \times 10^{-5}$	$+$	$i1.0000017386494497$
λ_{10}	$-3.16622826310105848 \times 10^{-5}$	$+$	$i1.0000008488497327$

Table 5.2: Computed roots near $i = \sqrt{-1}$ for a millionth degree random polynomial.

Table 5.2 shows the estimates of the 10 roots nearest $i = \sqrt{-1}$ that were computed using our implementation.

Table 5.3 shows several residuals for the computed roots. The value $\frac{|p(\lambda)|}{|\lambda p'(\lambda)|}$ comes

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

λ	$\kappa(\lambda)$	$\frac{\lvert p(\lambda)\rvert}{\lvert\lambda p'(\lambda)\rvert}$	$\frac{\lvert\lvert Cv-\lambda v\rvert\rvert_\infty}{\lvert\lvert C\rvert\rvert_\infty \lvert\lvert v\rvert\rvert_\infty}$
λ_1	1.3×10^0	1.9×10^{-17}	0
λ_2	1.7×10^0	8.8×10^{-18}	1.3×10^{-15}
λ_3	3.4×10^0	3.0×10^{-17}	4.3×10^{-16}
λ_4	1.5×10^0	8.3×10^{-17}	0
λ_5	1.4×10^0	4.7×10^{-16}	0
λ_6	2.3×10^0	1.1×10^{-16}	0
λ_7	3.3×10^0	3.2×10^{-17}	5.0×10^{-16}
λ_8	1.2×10^0	2.0×10^{-17}	0
λ_9	1.3×10^0	1.9×10^{-17}	0
λ_{10}	1.5×10^0	5.4×10^{-17}	0

Table 5.3: Error table for computed roots of a millionth degree random polynomial.

from the first order Taylor expansion of $p(\mu)$, where μ is the exact root, centered at the computed root λ,

$$p(\mu) = p(\lambda) + p'(\lambda)(\mu - \lambda) + O\left((\mu - \lambda)^2\right).$$

This implies that

$$|\mu - \lambda| = \frac{|p(\lambda)|}{|p'(\lambda)|} + O\left(|\mu - \lambda|^2\right).$$

If $|\mu - \lambda|$ is small then $\frac{|p(\lambda)|}{|p'(\lambda)|}$ is a good error estimate. To get a relative error estimate simply divide by the magnitude of λ, $\frac{|p(\lambda)|}{|\lambda p'(\lambda)|}$. The last column of Table 5.3 is the infinity norm residual of the Ritz pair (λ, v).

6. Conclusions. We have shown that arbitrary shifted, Fiedler companion matrices can be factored into $2n - 1$ essentially 2×2 matrices and therefore require $O(n)$ storage. This factorization gives immediate access to the factorization of the inverse of the shifted companion matrix as well. We have also shown that the cost of applying either of these factorizations to a vector is $O(n)$ making it possible to employ any shift and invert eigenvalue method efficiently.

Acknowledgment. I thank the editor and referee for their careful reading of the manuscript and for their questions, criticisms, and suggestions, which lead to significant improvements in the paper.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 26, pp. 231-247, May 2013

ELA

REFERENCES

[1] M. Fiedler. A note on companion matrices. *Linear Algebra and its Applications*, 372:325–331, 2003.

[2] J.L. Aurentz, R. Vandebril, and D.S. Watkins. Fast computation of the zeros of a polynomial via factorization of the companion matrix. *SIAM Journal on Scientific Computing*, 35(1):255–269, 2013.

[3] D.S. Watkins. *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods.* SIAM, Philadelphia, 2007.

[4] J.K. Cullum and R.A. Willoughby. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. 1 Theory.* Birkhauser, Boston, 1985.