All you should know about your “Companion”

Robert E Hartwig

Department of Mathematics
North Carolina State University
http://www4.ncsu.edu/~hartwig

Every mathematician will meet a good dose of linear algebra in his/her battle to reach nirvanna, be it as stepping stone to infinite dimensions, or as an entry into more abstract algebra, or as a computational tool in numerical analysis.

In this article we examine the story of the “companion matrix”

$$L[f(x)] \quad = \quad \begin{bmatrix} 0 & \cdots & & -f_0 \\ 1 & & & -f_1 \\ & \ddots & & \\ 0 & & 1 & -f_{n-1} \end{bmatrix},$$

which is associated with the monic polynomial $f(x) = f_0 + f_1x + \cdots + f_nx^n$ with $f_n = 1$, over a field $\mathbb{F}$.

These matrices are the “molecules” that lie at the heart of the whole field of linear algebra and its many applications, ranging from Canonical Forms to Systems Theory, Digital Image Processing and Numerical Analysis.

Their membership includes the “least periodic” matrix in the form of a nilpotent Jordan block, as well as the “most periodic” matrix, which undoubtedly is the circulant matrix.

Companion matrices really represent polynomials and appear whenever polynomials are involved. Add to this that polynomials are one of the most important concept in applicable mathematics, and so there is some ground for examining them.

They are in fact a realization of “finiteness”. Indeed, finiteness implies periodicity, periodicity implies the existence of annihilating polynomials and these in turn imply the existence of companion matrices.

Companion matrices and annihilating polynomials are two concepts which like “foot soldiers” are always “there” in the background! In fact, as in real life, it is the behavior of these “molecules” that dictates the behavior of matrices in general.

There are numerous reasons why these matrices play such a dominant role, and we shall not attempt to give a complete “all you should know about your companion” presentation.

In this note we shall demonstrate that many of the useful results involving companion matrices are a consequence of the *companion shift* property. We shall go through several of these and develop the notation as we go along. Since there are so many areas of application, some secrets of our companion will be left to the literature.

Before we present the shift condition, we mention some of its well known properties.

- Its transpose L^T is similar to L, which translates into the fact that any matrix is similar to its transpose.
- $f(x)$ is a (left) annihilating polynomial for $L(f)$. Indeed $f_\ell(L_f) = L^n + L^{n-1}f_{n-1} + \cdots + If_0 = 0$, which ensures that any matrix has an annihilating polynomial.
- Companion matrices are irreducible Hessenberg matrices that represent polynomials.
- They are the simplest matrices that one can write down with a prescribed characteristic polynomial. As such they are crucial building blocks in the construction of Canonical Forms.
- They are *non-derogatory*, i.e. the characteristic and minimal polynomials are equal. (blame Sylvester!)
- They are sparse (= most of their entries are zero).
- They are closely associated with **cyclic** subspaces, which are the most important subspaces of linear algebra, module theory, etc. Indeed, their use ranges from Canonical Forms to chain computations, as found in coding and numerical analysis (GMRES).
- The link to the cyclic subspaces is provided by the cyclic *chains*, which play a key role in the question of minimal polynomials and basis changes.

- Companion matrices induce a *companion shift*, on a chain of matrices, which can be manipulated in numerous ways. These shifts may be compensated for when the terms in the string satisfy suitable recurrence relations. It is this cancelation phenomena that is at the bottom of the some of the deeper theorems.
- Cyclic subspaces parallel cyclic groups, which in turn are fundamental building blocks in all of group theory.
- The companion matrix of $L(x^n - 1)$ is special. It presents itself when dealing with permutation matrices and is the linch pin in the whole field of Discrete Fourier Transform. Indeed, the roots of its polynomial generate the "mother of all groups", i.e. the group of n-th roots of unity.

Before we enter the realm of the companion matrix let us first clear up some more of the needed definitions and notations. Throughout this article all matrices will be over a field $\mathbb{F}$, but many results extend to the non-commutative (block) case.

Let $A \in \mathbb{F}_{n\times n}$ and $\mathbf{x} \in V = \mathbb{F}^n$. The characteristic and minimal polynomials of A are denoted by $\Delta_A(x)$ and $\psi_A(x) = \psi_V(x)$, respectively. The minimal annihilating polynomial (m.a.p.) of $\mathbf{x}$ relative to A is the monic polynomial $\psi_x(\lambda)$, of least degree, such that $\psi_x(A)\mathbf{x} = \mathbf{0}$. A Polynomial will be denoted by $p(x)$ or by $p(\lambda)$, when there is no ambiguity, and we use $\partial(.)$ for degree. The reciprocal polynomial of $f(x)$ is given by $\tilde{f}(x) = x^n f(1/x)$. We shall freely interchange L_f and $L(f)$ and will suppress the subscript where convenient. We use rk(.) and $\nu(.)$ for r rank and nullity and denote the Kronecker product by $\otimes$. A is regular if $AA^-A = A$ for some A^- and col$[\mathbf{x_1},..,\mathbf{x_n}]$ stands for $[\mathbf{x_1^T},..,\mathbf{x_n^T}]^\mathbf{T}$.

Besides $L(f)$, there are several other matrices that are also determined by $f(x)$. In particular we need its $n \times n$ symmetric Hankel matrix $G = G[f(x)]$, and the $(n+t)\times t$ basic *shift* matrix $S_t(f)$, generated by $f(x)$, which are given by

$$G_f = \begin{bmatrix} f_1 & f_2 & \cdots & & f_n \\ f_2 & f_3 & \cdots & f_n & 0 \\ \vdots & & & & \\ f_{n-1} & f_n & 0 & \cdots & 0 \\ f_n & 0 & & \cdots & 0 \end{bmatrix}, \text{ and}$$

$$S_t(f) = \begin{bmatrix} f_0 & & & 0 \\ f_1 & f_0 & & \\ \vdots & & \ddots & \\ f_n & f_{n-1} & \cdots & f_0 \\ 0 & f_n & & \vdots \\ \vdots & & \ddots & \\ 0 & & & f_n \end{bmatrix}.$$

For later use, we denote the "flip matrix" $G(x^n)$ by F. In many settings it is essential to use the transposed companion matrix L^T, rather than L – as for example with eigenvectors – because left and right multiplication are different. The fundamental relation between L and L^T is given by

$$L_f G_f = (L_f G_f)^T = G_f L_f^T \text{ or } G^{-1}LG = L^T,$$

and even holds in the non-commutative case. This identity is the reason why G is often referred to as the "intertwining" matrix or symmetrizer.

Consider L_f, where $f(x) = f_0 + f_1 x + \cdots + x^n$. We associate the coefficient vectors $\mathbf{f} = [f_0, \ldots, f_{n-1}]^T$ and $\hat{f} = [f_0, \ldots, f_n]^T$ and the chain matrices $X'_n = [1, x, ., x^{n-1}]$ and $Y_n = \begin{bmatrix} 1 \\ y \\ \vdots \\ y^{n-1} \end{bmatrix}$. The companion shift takes the form

$$(i)\ xX'_n - X'_n L_f = f(x)\mathbf{e}_n^T \ \ or\ \ (ii)\ L_f^T X_n - xX_n = -f(x)\mathbf{e}_n \tag{0.1}$$

which is trivial to verify and yet is the most important property of the companion matrix!

The basic shift matrix $S_m(f)$ has been introduced to take care of polynomial multiplication, i.e. of convolution. Again, let $f(x) = f_0 + f_1 x + \cdots + f_n x^n, g(x) = g_0 + g_1 x + \cdots + g_m x^m$, and $h(x) = f(x)g(x) = h_0 + h_1 x + \cdots + h_{m+n}x^{m+n}$, with coordinate columns $\bar{\mathbf{f}}, \bar{\mathbf{g}}$ and $\bar{\mathbf{h}}$. The two key results that we need are

$$\begin{aligned} f(x)X'_n &= X'_{2n} S_n(f) \\ S_{m+1}(f)\bar{\mathbf{g}} &= \bar{\mathbf{h}}. \end{aligned} \tag{0.2}$$

The latter reflects the convolution product $h_k = f_k g_0 + f_{k-1}g_1 + \cdots + f_0 g_k$.

There are numerous types of operations that we can now perform on this shift equation.

(i) We can *multiply* through by a suitable matrix.

(ii) We can *embed* this shift into a unimodular polynomial matrix – which will enable us show the equivalence of $xL - L$ to its Smith Normal Form diag$(f(x), I)$.

(iii) We can consider it as a *matrix equation* of the form $AX - XB$ =C, and use the *telescoping trick*, – i.e. repeatedly pre-multiplying by A and post multiplying by B, and add – to arrive at

$$A^k X - XB^k = \Gamma_k = A^{k-1}C + A^{k-2}CB + \cdots + CB^{k-1} \tag{0.3}$$

(iv) We may (formally) *differentiate* to give

$$X'^{(k)}_n (L_f - xI) = kX'^{(k-1)}_n - f^{(k)}(x)\mathbf{e}_n^T$$

(v) We may *evaluate* the identity at $x = a$ or replace x by a matrix B, giving us useful block identities.

(vi) We may *combine* any of the above such as the companion shift with the basic shift $S_n(g)$, or differentiation followed by multiplication.

Let us now present each of the above and see how this application can be used.

1 Multiplication

If we multiply (0.1) by G, we meet the family of *adjoint* polynomials $f_i(x)$ given by

$$F'_n = [f_0(x), f_1(x), \dots, f_{n-1}(x))] = [1, x, x^2, \dots, x^{n-1}]G_f,$$

or in detail

$$f_k(x) = f_{k+1} + f_{k+2}x + \cdots + f_n x^{n-k-1},\ k = 0, 1, \dots, n-1.$$

It should be noted that $f_{-1}(\lambda) = f(\lambda)$ and that $f_{n-1}(\lambda) = f_n = 1$.

If we multiply (0.1)-(i) on the right by G we obtain the *adjoint shift condition*

$$xF'_n - F'_n L_f^T = f(x)\mathbf{e}_1^T, \tag{1.4}$$

which is equivalent to the recurrence relation $f_{k-1}(x) = f_k + xf_k(x)$, $k = 0, 1, \dots, n-1$.

If we right multiply (0.1) by $\mathrm{adj}(xI - L)$ then we generate $f(x)\mathbf{e}_n^T \mathrm{adj}(xI - L) = X_n^T(xI - L)\mathrm{adj}(xI - L) = X_n^T f(x)I$, from which we may cancel $f(x)$ to give

$$X'_n = [1, x, ., x^{n-1}] = \mathbf{e}_n^T[\mathrm{adj}(xI - L_f)].$$

2 Substitution

Given a matrix $A \in \mathbb{F}_{n\times n}$, if we replace x by A in (0.1) and (1.4) we see that

$$A[I, A, \dots, A^{m-1}] = \\ [I, A, \dots, A^{m-1}][L_A(f) \otimes I] + [0, \dots, 0, f(A)]. \tag{2.5}$$

and

$$A[A_0, A_1, \dots, A_{m-1}] = \\ [A_0, A_1, \dots, A_{m-1}][L_A^T(f) \otimes I] + [f(A), 0, \dots, 0].$$

If in addition $f(x) = \Delta_A(x)$ and $f(A) = 0$, then we obtain the coefficients $A_i = f_i(A)$ in the expansion $\mathrm{adj}(xI - A) = \sum_{i=0}^{n-1} A_i x^i$. Not surprisingly we can use the companion shift to actually characterization a companion matrix. Indeed,

Proposition 2.1. *Given a matrix B with minimal polynomial $p(x)$ of degree n. An $n \times n$ matrix X equals $L(p)$ iff*

$$B[I, B, \dots, B^{n-1}] = [I, B, \dots, B^{n-1}](X \otimes I). \tag{2.6}$$

Proof. If $X = L(p)$ then take $f = p$ in (2.5). Conversely, if (2.6) holds we select $f = p$ in (2.5). Subtracting shows that $[I, B, \dots, B^{n-1}][(L(p) - X) \otimes I] = 0$. Since the powers in the chain are independent it follows that $X = L(p)$. □

We may now introduce a second matrix B, and multiply (2.5) through by $(I \otimes B)$ to give for any monic polynomial $f(x)$

$$A[B, AB, \dots, A^{r-1}B] = \\ [B, AB, \dots, A^{r-1}B](L_f \otimes I) + [0, \dots, 0, f(A)B].$$

Chains of the form $[B, AB, \dots, A^{r-1}B]$ are of considerable importance in linear control and systems theory. We shall mainly focus on the case where B is a column $\mathbf{x}$, in which case the chain matrix $K_r(\mathbf{x}, A) = [\mathbf{x}, A\mathbf{x}, A^2\mathbf{x}, \dots, A^{r-1}\mathbf{x}]$ is referred to as a Krylov matrix. There are now two cases of interest.

(i) Suppose that the m.a.p of ψ_x has degree $\partial(\psi_x) = r$. Then $A^r\mathbf{x}$ is the smallest (= first) power that is a linear combination of the previous powers, and as such the r links in the chain matrix $K_r(\mathbf{x}\ , A) = [\mathbf{x}, A\mathbf{x}, A^2\mathbf{x}, \dots, A^{r-1}\mathbf{x}]$ are linearly independent and

$$AK_r(\mathbf{x}, A) = K_r(\mathbf{x}, A)L[\psi_x(\lambda)].$$

Completing the matrix $K_r(\mathbf{x}, A)$ to an invertible matrix $Q = [K_r, B]$, we then arrive at

$$AQ = Q\left[\begin{array}{c|c} L(\psi_x) & E \\ \hline 0 & D \end{array}\right] \text{ i.e. } Q^{-1}AQ = \left[\begin{array}{c|c} L(\psi_x) & E \\ \hline 0 & D \end{array}\right].$$

This will shortly be used as the first step in the derivation of the Cyclic-Decomposition Theorem.

(ii) If, on the other hand, we take the first n links in the chain we obtain $K_n(\mathbf{x}, A) = [\mathbf{x}, A\mathbf{x}, A^2\mathbf{x}, \dots, A^{n-1}\mathbf{x}]$. Now $A^n\mathbf{x}$ must be a linear combination of lower powers, say $A^n\mathbf{x} = -[(f_0\mathbf{x} + f_1A\mathbf{x} + \cdots + f_{n-1}A^{n-1}\mathbf{x}]$. We see that

$$AK_n = K_nL(f),$$

where $f(x) = f_0 + f_1x + \cdots + x^n$. It is clear that $K_n(\mathbf{x}, A)$ will be non-singular iff $\partial(\psi_x) = n$, in which case A is non-derogatory and $A = K_nL(f)K_n^{-1}$.

Associated with the above chain is the cyclic subspace $Z_{\mathbf{x}}(A) =< \mathbf{x}, A\mathbf{x}, A^2\mathbf{x}, \dots, >$ generated by $\mathbf{x}$. It is also referred to as the Krylov subspace generated by x, and is used, for example, in the GMRES method of numerical analysis.

A vectorspace V is called cyclic if $V = Z_{\mathbf{u}}(A)$ for some vector $\mathbf{u}$ in V. For the case where $V = \mathbb{F}^n$ this means that $V =< \mathbf{u}, A\mathbf{u}, \dots, A^{n-1}\mathbf{u} >= R(K_n(\mathbf{u}, A))$.

Cyclic subspaces parallel the concept of cyclic groups, which are by far the most important type of group. The following is, for example, the analog to the fundamental theorem of finite cyclic subgroups.

Proposition 2.2. *If $V = Z_u(A)$ and W is an A-invariant subspace then*

(i) W is also a cyclic (sub)space.

ii) $W = Z_{g(A)\mathbf{u}}$ for some polynomial $g(\lambda)$.

It is easily verified that the vector $\mathbf{y} = g(A)\mathbf{u}$ indeed has a m.a.p equal to $\psi_y = \psi_A/g$.

As such, it should come as no surprise that cyclic subspaces also play an important role in several areas of applied linear algebra such as in coding and in linear control and *pole placement.* Indeed, cyclic codes contain the BCH codes, which are one of the most important families of error-correcting codes. The key question: When is a vectorspace V cyclic? is really a question about companion matrices

3 Embedding

Next, we embed X'_n into a unimodular matrix

$$K(x) = \left[\begin{array}{c|cccc} 1 & x & x^2 & \dots & x^{n-1} \\ \hline 0 & 1 & x & & x^{n-2} \\ & & 1 & & \vdots \\ & & & \ddots & x \\ 0 & & & & 1 \end{array}\right],$$

and then compute $K(x)[xI - L(f)] = \begin{bmatrix} \mathbf{0}^T & f(x) \\ -I & \mathbf{u} \end{bmatrix}$, where $\mathbf{u}^T = [f_0(x), f_1(x), \dots, f_{n-2}(x)]^T$. Selecting $R(x) = \begin{bmatrix} \mathbf{u} & -I \\ 1 & 0 \end{bmatrix}$, we see that

$$K(x)[xI - L(f)]R(x) = \begin{bmatrix} f(x) & 0 \\ 0 & I \end{bmatrix}.$$

Since $K(x)$ and $R(x)$ are unimodular with $K(x)^{-1} = I - xJ_n(0)$, we have obtained the Smith Normal Form $\text{diag}(f(x), I)$ of $L(f)$.

Taking determinants shows further that $\Delta_L(x) = f(x)$ so that L_f is indeed non-drogatory.

4 Some basic Identities

Before we progress, we shall need several basic identities that illustrate how companion matrices deal with polynomial properties. The key shift property of L comes from the following.

Proposition 4.1. *If $L = L_f$ and $\mathbf{f} = [f_0, \dots, f_{n-1}]^T$, then*

(i) $L^i\mathbf{e}_1 = \mathbf{e}_{i+1}$ for $i = 0, 1, \dots, n-1$ and $L^n\mathbf{e}_1 = \mathbf{f}$.

(ii) $I = [\mathbf{e}_1, L\mathbf{e}_1, \dots, L\mathbf{e}_{n-1}] = [\mathbf{e}_1, L\mathbf{e}_1, \dots, L^{n-1}\mathbf{e}_1]$.

(iii) $L^k = [\mathbf{e}_{k+1}, L\mathbf{e}_{k+1}, \dots, L^{n-1}\mathbf{e}_{k+1}]$, for $k = 1, 2, \dots$

From these we obtain the curious by-product that

$$col([I, L, L^2, \dots, L^{n-1}]) = col(\begin{bmatrix} I \\ L \\ \vdots \\ L^{n-1} \end{bmatrix}).$$

We next show that polynomials in L and L^T generate Krylov chains and that they are completely determined by their first and last columns respectively.

Lemma 4.2. *Let $L = L(f)$, $g(x) = \sum_{i=0}^n g_i x^i$ and $h(x) = \sum_{i=0}^k h_i x^i$, with $k < n$, $h_k \neq 0$ and associated vectors $\mathbf{g} = [g_0, g_1, \dots, g_{n-1}]^T$, $\mathbf{h} = [h_0, h_1, \dots, h_{n-1}]^T$ and $\boldsymbol{\gamma} = \mathbf{g} - g_n\mathbf{f}$. Then*

(i) $g[L(f)] = [\boldsymbol{\gamma}, L\boldsymbol{\gamma}, \dots, L^{n-1}\boldsymbol{\gamma}]$.

(ii) $\text{rk}[h(L)] \geq n - k$ with equality if $h|f$.

(iii) $g(L)G_f = [f_0(L_f)\boldsymbol{\gamma}, \dots, f_{n-1}(L_f)\boldsymbol{\gamma}]$.

(iv) $g[L(f)]\mathbf{h} = h[L(f)]\boldsymbol{\gamma}$.

(v) $g(L^T)\mathbf{e}_j = f_{j-1}(L^T)g(L^T)\mathbf{e}_n$.

Proof. (i) $g(L)\mathbf{e}_1 = \sum_{i=0}^{n-1}(g_i - g_n f_i)L^i\mathbf{e}_1 = \sum_{i=0}^{n-1}(g_i - g_n f_i)\mathbf{e}_{i+1} = \boldsymbol{\gamma}$.
(ii) Matrix $[\mathbf{h}, \dots, L^{n-k-1}\mathbf{h}] = S_{n-k}(h)$ has rank $n-k$. If $f = hq$, then $\partial(q) = n-k$ and $\text{rk}[q(L)] \geq n-(n-k)$. Also $0 = h(L)q(L)$, which shows that $\nu[h(L)] \geq k$.
(iii) $[f_0(L_f)\boldsymbol{\gamma}, \dots, f_{n-1}(L_f)\boldsymbol{\gamma}] = [I, L, \dots, L^{n-1}](G_f \otimes I)(I \otimes \boldsymbol{\gamma}) = [I, L, \dots, L^{n-1}](I \otimes \boldsymbol{\gamma})G_f = [\boldsymbol{\gamma}, L\boldsymbol{\gamma}, \dots, L^{n-1}\boldsymbol{\gamma}]G = g(L)G$.
(iv)$[\boldsymbol{\gamma}, L\boldsymbol{\gamma}, .., L^{n-1}\boldsymbol{\gamma}]\mathbf{h} = \mathbf{h}(\mathbf{L})\boldsymbol{\gamma}$.
(v) $g(L^T)\mathbf{e}_j = G^{-1}[g(L)G]\mathbf{e}_j = G^{-1}f_{j-1}(L)\boldsymbol{\gamma} = [G^{-1}f_{j-1}(L)]g(L)\mathbf{e}_1 = [G^{-1}f_{j-1}(L)]g(L)G\mathbf{e}_n = f_{j-1}(L^T)g(L^T)\mathbf{e}_n$. □

Part one shows that the map must have degree n. Part two shows that if $d = (f, g)$ then $\text{rk}[g(L_f)] = \text{rk}[d(L_f)] = n - \partial(d)$, which illustrates the close connection between gcds and companion matrices, and is crucial in systems theory.
Next we observe that G does symmetrize all powers of L, i.e. $L_f^k G_f$ is symmetric. In fact it follows by induction that, for $k = 1, 2, \dots$,

$$L_f^k G_f = \text{diag}(-\begin{bmatrix} 0 & & f_0 \\ & \iddots & \vdots \\ f_0 & \cdots & f_{k-1} \end{bmatrix}, \begin{bmatrix} f_{k+1} & \cdots & f_n \\ \vdots & \iddots & \\ f_n & & 0 \end{bmatrix}).$$

We now come to a couple of useful inverses.

The inverse of $L(f)$ exists exactly when f_0 is invertible and has the form of flipped companion matrix, i.e.

$$L(f)^{-1} = -[I(f_1 f_0^{-1}) + L(f_2 f_0^{-1}) + \cdots + L^n(f_0^{-1}].$$

The inverse of G on the other hand, is again a Hankel matrix. Indeed,

$$\begin{aligned} G[\mathbf{e}_n, L^T\mathbf{e}_n, \ldots, (L^T)^{n-1}\mathbf{e}_n] &= [G\mathbf{e}_n, GL^T\mathbf{e}_n, \ldots, G(L^T)^{n-1}\mathbf{e}_n] \\ &= [\mathbf{e}_1, LG\mathbf{e}_n, \ldots, L^{n-1}G\mathbf{e}_n] \\ &= [\mathbf{e}_1, L\mathbf{e}_1, \ldots, L^{n-1}\mathbf{e}_1] = I. \end{aligned}$$

Transposing now gives

$$G^{-1} = [\mathbf{e}_n, L^T\mathbf{e}_n, \ldots, (L^T)^{n-1}\mathbf{e}_n] = \begin{bmatrix} \mathbf{e}_n^T \\ \mathbf{e}_n^T L \\ \mathbf{e}_n^T L^2 \\ \vdots \\ \mathbf{e}_n^T L^{n-1} \end{bmatrix}.$$

Using this in turn yields

$$I = GG^{-1} = G \begin{bmatrix} \mathbf{e}_n^T \\ \mathbf{e}_n^T L \\ \mathbf{e}_n^T L^2 \\ \vdots \\ \mathbf{e}_n^T L^{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_n^T f_0(L) \\ \vdots \\ \mathbf{e}_n^T f_{n-1}(L) \end{bmatrix},$$

establishing that $\mathbf{e}_n^T f_i(L) = \mathbf{e}_{i+1}^T$. We conclude this section with some of the interaction between $L(f)$ and the matrix $N = E_{1,n}$ and assume that $\partial(g) < n$.

Proposition 4.3. *The following hold:*

(i) $L[f(x) + g(x)] = L(f) - \mathbf{e}_n \mathbf{g}^T$.

(ii) $L[f(x) - 1] = L + N$ and $f(L + N) = I$.

(iii) $L[f(x) - x^k] = L + E_{k+1,n}$ and $f(L + E_{k+1,n}) = (L + E_{k+1,n})^k$.

(iv) $NL_f^k N = 0$ for $k = 0, 1, \ldots, n-1$ and $NL^{n-1}N = N$.

(v) $g(L_f + N)N = g(L_f)N + g_{n-1}N$.

(vi) $f(L_f + N)N = N = Nf(L_f + N)$.

(vii) $(L + N)^r - L^r = \Gamma_r(L, N, L) = L^{r-1}N + L^{n-2}NL + \cdots + NL^{r-1}$, $r = 1, \ldots, n-2$.

Proof. (ii)-(iii) $f(x) - x^k$ is an ap for $L + E_{k+1,n}$. (iv) Follows from (4.1)-(i). (v) $g(L+N)\mathbf{e}_1 = g[L(f-1)]\mathbf{e}_1 = L(f-1)\mathbf{g} = (L+N)\mathbf{g} = L\mathbf{g} + N\mathbf{g} = g(L)\mathbf{e}_1 + g_{n-1}\mathbf{e}_1$. (vi) $N\Gamma_r = 0$ for $r = 0, \ldots, n-1$. □

5 Corner matrices

Besides multiplication or evaluation there are two other operations that we can apply to the companion shift, and these are telescoping or differentiation. Actually the corner matrix acts very much like "differentiation". For example, for any polynomial $g(x)$,

$$\begin{aligned} g(\begin{bmatrix} x & 1 \\ 0 & x \end{bmatrix}) &= \begin{bmatrix} g(x) & g'(x) \\ 0 & g(x) \end{bmatrix} \\ &= g(\begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix}) + g'(x) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

More generally, the corner matrices Γ_k appear in the powers of $\begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$. The difference form now becomes $\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}^k - \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}^k = \begin{bmatrix} 0 & \Gamma_k \\ 0 & 0 \end{bmatrix}$, where Γ_k satisfies the down-shift recurrence relation

$$\Gamma_{k+1}(A, C, D) = A\Gamma_k + CD^k = A^k C + \Gamma_k D.$$

If we now have a second polynomial $g(x) = g_0 + g_1 x + \cdots + x^N$ then $g(M) = \begin{bmatrix} g(A) & \Gamma_g A, C, B) \\ 0 & g(D) \end{bmatrix}$, where

$$\begin{aligned} \Gamma_g &= \sum_{i=1}^{N} g_i \Gamma_i = \sum_{i=0}^{n-1} g_i(A) C B^i \\ &= [I, A, \ldots, A^{n-1}][G_g \otimes C] \begin{bmatrix} I \\ B \\ \vdots \\ B^{n-1} \end{bmatrix} \qquad (5.7) \end{aligned}$$

and the $g_i(x)$ are the adjoint polynomials affiliated with $g(x)$. The difference now becomes $g(\begin{bmatrix} A & C \\ 0 & D \end{bmatrix}) - g(\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}) = \sum_{i=1}^{N} \begin{bmatrix} f_i(A) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & C \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & B^i \end{bmatrix}$.

6 Companion matrices, Resultants and Bezoutians

We now present an example in which we telescope the companion shift equation resulting in a simple relation between companion matrices, resultants and Bezoutians. The latter enters the realm of root location, stability analysis, and gcd-degree computation.

Given two monic polynomials $f(x)$ and $g(x)$ of degree n. The bilinear form associated with $f(x)$ is the difference quotient

$$\frac{f(x) - f(y)}{x - y} = X_n' G_f Y_n = \sum_{i=0}^{n-1} f_i \Gamma_i(x, y).$$

With the polynomial pair $f(x)$, $g(x)$ we may associate, the form

$$\mathbb{B}(f, g) = \frac{f(x)g(y) - f(y)g(x)}{x - y} = \sum_{i,j=0}^{n-1} b_{ij} x^i y^j = X_n' B Y_n.$$

It is bilinear and anti-symmetric, i.e. $\mathbb{B}(g,f) = -\mathbb{B}(f,g)$. The $n \times n$ matrix $B = \mathbb{B}(f,g)$ is called the Bezoutian (of Hankel type), and is symmetric. It is clear that

$$\begin{aligned}\mathbb{B}(f,g) &= g(x)\frac{f(x)-f(y)}{x-y} - f(x)\frac{g(x)-g(y)}{x-y} \\ &= g(x)[X_n' G_f Y_n] - f(x)[X_n' G_g Y_n]\end{aligned}$$

In order to tackle the bilinear form $[g(x)X_n']G_f Y_n$ we use the basic shift on $g(x)X_n'$ to simplify the result. Since $t = n$, it is convenient to split $S_n(f) = \begin{bmatrix} S_n^-(f) \\ S_n^+(f) \end{bmatrix}$, where

$$S_n^-(f) = \begin{bmatrix} f_0 & & 0 \\ \vdots & \ddots & \\ f_{n-1} & \cdots & f_0 \end{bmatrix},\ S_n^+(f) = \begin{bmatrix} f_n & \cdots & f_1 \\ & \ddots & \\ 0 & & f_n \end{bmatrix}.$$

Because of their Toeplitz structure, representing polynomial multiplication, we know that $S_n^-(f)$ and $S_n^-(g)$ commute, as do their "plus" counter parts. We can now split the basic shift (0.2) into

$$f(x)X_n' = X_{2n}' S_n(f) = X_n' S_n(f)^- + x^n X_n' S_n(f)^+. \quad (6.8)$$

Next we recall the companion shift (0.1), which has the form $AX - XB = C$, where $A = x$, $B = L(f)$, $X = X_n'$ and $C = f(x)\mathbf{e}_n^T$. Telescoping as in (0.3) we obtain

$$x^k X_n' - X_n' L^k = \Gamma_k(xI, f(x)\mathbf{e}_n^T, L) = f(x)\mathbf{e}_n^T \Gamma_k(xI, L) \quad (6.9)$$

If the second polynomial is given by $g(x) = g_0 + g_1 x + \cdots + x^n$, then we pre-multiply (6.9) by g_k and sum, giving

$$g(x)X_n' - X_n' g(L_f) = f(x)\mathbf{e}_n^T \Gamma_g(xI, L) = f(x)\mathbf{q}^T(x), \quad (6.10)$$

We next use the adjoint polynomials in

$$\begin{aligned}\mathbf{q}^T &= \mathbf{e}_n^T \Gamma_g = \sum_{i=0}^{n-1} g_i(x)\mathbf{e}_n^T L_f^i \\ &= [g_0(x), \ldots, g_{n-1}(x)] \begin{bmatrix} \mathbf{e}_n^T \\ \mathbf{e}_n^T L_f \\ \vdots \\ \mathbf{e}_n^T L_f^{n-1} \end{bmatrix} \\ &= X_n' G_g G_f^{-1} = X_n' Q, \quad \text{where } Q = (G_g G_f^{-1}).\end{aligned}$$

Applying the basic shift in (6.10) now gives

$$X_{2n}' S_n(g) - X_{2n}' \begin{bmatrix} g(L_f) \\ 0 \end{bmatrix} = X_{2n}' S_n(f) Q,$$

from which we obtain the identity

$$S_n(g)G_f - \begin{bmatrix} g(L_f) \\ 0 \end{bmatrix} G_f = S_n(f)G_g.$$

We could also have telescoped the adjoint shift equation. Splitting this then produces

$$\begin{bmatrix} S_n^-(g)G_f \\ S_n^+(g)G_f \end{bmatrix} - \begin{bmatrix} S_n^-(f)G_g \\ S_n^+(f)G_g \end{bmatrix} = \begin{bmatrix} g(L_f)G_f \\ 0 \end{bmatrix}$$

in which we equate blocks to yields

$$S_n^-(g)G_f - S_n^-(f)G_g = g(L_f)G_f \text{ and } S_n^+(g)G_f = S_n^+(f)G_g) \quad (6.11)$$

The latter follows from the fact that $S_n^+(f)$ and $S_n^+(g)$ commute and

$$S_n^+(f)F = G_f,\ FG_f = S_n^-(\tilde{f}),\ \ S_n^+(f)^T = S_n^-(\tilde{f}), \quad (6.12)$$

where $\tilde{f}$ is the reciprocal polynomial. Using the split basic shift (6.8) we see that

$$[g(x)X_n']G_f Y_n = X_n' S_n^-(g)G_f Y_n + x^n X_n' S_n^+(g)G_f Y_n$$

while

$$[f(x)X_n']G_g Y_n = X_n' S_n^-(f)G_g Y_n + x^n X_n' S_n^+(f)G_g Y_n\ .$$

Subtracting them, we obtain

$$\begin{aligned}\mathbb{B}(f,g) = &X_n'[S_n^-(g)G_f - S_n^-(f)G_g]Y_n \\ &+ x^n X_n'[S_n^+(g)G_f - S_n^+(f)G_g]Y_n,\end{aligned}$$

in which the second term vanished because of (6.11). Extracting the matrix we see that $B(f,g) = S_n^-(g)G_f - S_n^-(f)G_g$ which on account of (6.11) gives Barnett's formula

$$B(f,g) = S_n^-(g)G_f - S_n^-(f)G_g = g(L_f)G_f.$$

Lastly we introduce the resultant matrix $M(g,f) = \begin{bmatrix} S_n^-(g) & S_n^-(f) \\ S_n^+(g) & S_n^+(f) \end{bmatrix}$ and the two row matrices

$$T = \begin{bmatrix} I & -S_n^-(f)[S_n^+(f)]^{-1} \\ 0 & I \end{bmatrix} \text{ and } \mathrm{U} = \begin{bmatrix} I & 0 \\ -Q & I \end{bmatrix}.$$

We subsequently compute the triplet TM(g,f)U in two ways, establishing that

$$\begin{aligned}\begin{bmatrix} \zeta & 0 \\ 0 & S_n^+(f) \end{bmatrix} &= \begin{bmatrix} \zeta & 0 \\ S_n^+(g) & S_n^+(f) \end{bmatrix} \begin{bmatrix} I & 0 \\ -Q & I \end{bmatrix} \\ &= \begin{bmatrix} I & -S_n^-(f)[S_n^+(f)]^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} g(L_f) & S_n^-(f) \\ 0 & S_n^+(f) \end{bmatrix} \\ &= \begin{bmatrix} g(L_f) & 0 \\ 0 & S_n^+(f) \end{bmatrix}.\end{aligned}$$

As such we see that

$$g(L_f) = \zeta = S_n^-(g) - S_n^-(f)[S_n^+(f)]^{-1} S_n^+(g)\ ,$$

which is the (2,2) Schur complement of M(g,f). In conclusion we may use (6.12) to compute

$$M(f,g)M^T(\tilde{g}, -\tilde{f}) = \begin{bmatrix} S_n^-(f) & S_n^-(g) \\ S_n^+(f) & S_n^+(g) \end{bmatrix} \begin{bmatrix} S_n^+(g) & S_n^-(g) \\ -S_n^+(f) & -S_n^-(f) \end{bmatrix},$$

which reduces to diag(K, N), where

$$K = -B(f,g)F \ \ \textit{and} \ \ N = -FB(\tilde{f}, \tilde{g}).$$

Needless to say, there are numerous generalizations of this concept to multivariate or non-commutative settings.

7 Generalized Adjoint chains

The idea of an adjoint chain may be extended by using a *block* Hankel matrix G. Indeed, suppose we are given the $m \times n$ polynomial matrix $\mathscr{F}(x) = F_0 + F_1 x + \cdots + F_N x^N$. The *adjoint polynomials* associated with this master polynomial are defined by

$$\mathscr{F}_k(x) = F_{k+1} + F_{k+2}x + \cdots + F_N x^{N-k-1}, k = 0, \ldots, N-1,$$

with in addition $\mathscr{F}_N(x) = 0$, $\mathscr{F}_{-1}(x) = \mathscr{F}(x)$ and $\mathscr{F}_{N-1}(x) = F_N$. As in the scalar case they can be expressed in block matrix form as

$$[\mathscr{F}_0(x),\ \mathscr{F}_1(x), \cdots, \mathscr{F}_{N-1}(x)] = [1, x, \cdots, x^{N-1}]G(\mathscr{F}),$$

where $G = G(\mathscr{F})$ is the block Hankel matrix

$$G = \begin{bmatrix} F_1 & F_2 & \ldots & & F_N \\ F_2 & F_3 & & F_N & 0 \\ \vdots & & & & \\ F_{n-1} & F_N & 0 & & \\ F_N & 0 & \ldots & & 0 \end{bmatrix}.$$

The key feature of these polynomials is that they satisfy the down-shift Recurrence Relation

$$x.\mathscr{F}_k(x) = \mathscr{F}_{k-1}(x) - F_k \quad k = 0, 1, \cdots, N-1,$$

in which we may replaced x by a suitable square matrix A on the left, or D, on the right.

Using the block recurrence we may write

$$\begin{bmatrix} \mathscr{F}_1(x) \\ \vdots \\ \mathscr{F}_N(x) \end{bmatrix} .x = \begin{bmatrix} \mathscr{F}_0(x) \\ \vdots \\ \mathscr{F}_{N-1}(x) \end{bmatrix} - \begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix}$$

and by using the companion structure we also have

$$(L_f \otimes I) \begin{bmatrix} \mathscr{F}_1(x) \\ \vdots \\ \mathscr{F}_N(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathscr{F}_1(x) \\ \vdots \\ \mathscr{F}_{N-1}(x) \end{bmatrix} - \begin{bmatrix} f_0 I \\ \vdots \\ f_{N-1} I \end{bmatrix} \mathscr{F}_N(x).$$

Subtracting these we arrive at the generalized adjoint-companion shift identity

$$(L_f \otimes I) \begin{bmatrix} \mathscr{F}_1(x) \\ \vdots \\ \mathscr{F}_N(x) \end{bmatrix} - \begin{bmatrix} \mathscr{F}_1(x) \\ \vdots \\ \mathscr{F}_N(x) \end{bmatrix} x = \begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix} - \begin{bmatrix} \mathscr{F}_0(x) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{7.13}$$

together with a row analog. It should be noted that the indices differ by one from those in (1.4), and it goes without saying that we may now again replace x by a suitable matrix D.

8 The Cyclic Decomposition Theorem

This theorem is a statement about the periodicity of finite dimensional objects, and is realized in terms of companion matrices.
There are essentially *two* versions of this theorem. A weak version, which is easier to prove, and a strong version, which requires much more firepower. The weak version says that any matrix A in $V = \mathbb{F}_{n \times n}$ is similar to a direct sum of companion matrices. Or equivalently, that any vectorspace over a field $\mathbb{F}$ can be decomposed as a direct sum of "cyclic" subspaces. It may be considered as a special case of the fundamental, theorem of abelian groups.

The best example is that of a permutation, which is a product of distinct cycles. In terms of matrices this says that any permutation matrix is similar to direct sum of matrices of the form $L(x^r - 1)$.

We shall use the adjoint-companion shift (7.13) to derive the "strong"' version of this theorem – which is often called the Rational Canonical Form – in which, in addition, the minimal polynomials of the companion matrices **interlace**. The proof is short and does not use quotient spaces.

When this theorem is combined with the Primary Decomposition Theorem, they will spawn the Jacobson and Jordan Canonical Forms.

Given a matrix A in $V = \mathbb{F}_{n \times n}$, with minimal polynomial ψ_A. Select a maximal vector $\mathbf{x}$, for which $\psi_{\mathbf{x}} = \psi_A = f(\lambda) = f_0 + f_1\lambda + \cdots + \lambda^m$. The existence of such a vector follows as for finite abelian groups G, when we replace the order $O(.)$ of an element by the m.a.p of a vector. Indeed in G, if $O(a) \nmid O(b)$ then there exists z in G such that $O(b)|O(z)$ but $b \neq z$ and if $O(y)$ is maximal then $O(a)|O(y)$ for all a in G.

We then form the chain matrix $K = [\mathbf{x}, A\mathbf{x}, \ldots, A^{m-1}\mathbf{x}]$, which has rank m, and complete it to a basis $Q = [K, B]$ for V. Then $AQ = Q \left[\begin{array}{c|c} L_f & C \\ \hline 0 & D \end{array}\right]$, for some C and D. Since Q is invertible, $Q^{-1}AQ = \left[\begin{array}{c|c} L & C \\ \hline 0 & D \end{array}\right] = M$, and thus $\psi_M = \psi_A = f$. It now follows that $0 = f(M) = \begin{bmatrix} f(L) & \Gamma_f \\ 0 & f(D) \end{bmatrix}$, in which the corner block takes the form

$$\Gamma_f = \sum_{k=0}^{m} f_k \sum_{j=0}^{k-1} L^{k-j-1} C D^j = \sum_{i=0}^{m-1} f_i(L) C D^i = 0 \tag{8.14}$$

and the $f_k(\lambda)$ are the usual adjoint polynomials of $f(x)$. Suppose now that $C = \begin{bmatrix} \gamma_1^T \\ \vdots \\ \gamma_m^T \end{bmatrix}$ and set $\mathscr{F}_k(x) =$

$\sum_{i=k}^{m-1} \gamma_{i+1}^T x^{i-k}$ and $\mathscr{F}_m(x) = 0$. It is easily seen that $\mathscr{F}_k(x)$ satisfies $\mathscr{F}_k(x).x = \mathscr{F}_{k-1}(x) - \gamma_k^T$ so that we can use the adjoint-shift identity (7.13)

$$L(f)\begin{bmatrix} \mathscr{F}_1(D) \\ \vdots \\ \mathscr{F}_m(D) \end{bmatrix} - \begin{bmatrix} \mathscr{F}_1(D) \\ \vdots \\ \mathscr{F}_m(D) \end{bmatrix} D = C - \begin{bmatrix} \mathscr{F}_0(D) \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

in which $\mathscr{F}_0(D) = \sum_{i=0}^{m-1} \gamma_{i+1}^T D^i = \sum_{i=0}^{m-1} \mathbf{e}_{i+1}^T C D^i$. Using the fact that $\mathbf{e}_m^T f_i(L) = \mathbf{e}_{i+1}^T$, this reduces to $\mathscr{F}_0(D) = \sum_{i=0}^{m-1} \mathbf{e}_m^T f_i(L) C D^i = \mathbf{e}_m^T \sum_{i=0}^{m-1} f_i(L) C D^i = \mathbf{e}_m^T \Gamma_f$, and thus, by (8.14), vanishes. In other words we have constructed a solution X to the matrix equation $L(f)X - XD = C$. This means that $\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}\left[\begin{array}{c|c} L(f) & C \\ \hline 0 & D \end{array}\right]\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} = \left[\begin{array}{c|c} L(f) & 0 \\ \hline 0 & D \end{array}\right]$, and consequently $A \approx M \approx \left[\begin{array}{c|c} L(f) & 0 \\ \hline 0 & D \end{array}\right]$, in which $\psi_D | \psi_M = f$. It goes without saying that we may repeat the above steps with D to obtain a direct sum decomposition

$$A \approx \mathrm{diag}[L(\psi_1), L(\psi_2), \ldots, L(\psi_t)],$$

where $\psi_t \mid \psi_{t-1} \mid \cdots \mid \psi_2 \mid \psi_1$.

The polynomials $\mathscr{J}_A = (\psi_1, \ldots, \psi_{t-1})$ are unique, and are called the **invariant factors** of A. They completely characterize similarity and do not depend on any possible factorization of polynomials, and were obtained by only using "rational operations". Hence the alternative name of "Rational Canonical Form".

The uniqueness of this Canonical Form follows at once, if we recall that $\psi_1 = \psi_A$ is unique and then apply the following elementary result.

Lemma 8.1. *If* $M = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \approx \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix} = N$ *then* $\psi_B = \psi_C$.

Proof. $\psi_B(M) = \begin{bmatrix} \psi_B(A) & 0 \\ 0 & 0 \end{bmatrix} \approx \begin{bmatrix} \psi_B(A) & 0 \\ 0 & \psi_B(C) \end{bmatrix}$. Taking ranks shows that $\mathrm{rk}[\psi_B(C)] = 0$ and thus $\psi_B(C) = 0$ and $\psi_C \mid \psi_B$. By symmetry, it also follows that $\psi_B \mid \psi_C$, ensuring equality. □

Now if $A \approx \mathrm{diag}[L(\psi), D] \approx \mathrm{diag}[L(\psi), E]$ then, by applying Lemma (8.1), we see that $\psi_D = \psi_E$, so that we can indeed continue the reduction process with D or with E. The same polynomials will be obtained.

It is of interest to note that the above method can actually also be used to prove that similarity of $\begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$ and $\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$, ensures that $AX - XD = C$ has a solution.

9 Differentiation

If the field is closed we may use the companion shift to obtain the Jordan form for $L(f)$, but since we use right eigenvectors, it is more convenient to use L_f^T. The transposed companion shift takes the form

$$L_f^T X_n(x) - X_n(x)x = -f(x)\mathbf{e}_n.$$

Now $f(a) = 0$ iff $X_n(a)$ is an eigenvector for L^T associated with eigenvalue a. The corresponding eigenvector for L_f will be $\mathscr{F}(a) = G_f X_n(a) = \begin{bmatrix} f_0(a) \\ \vdots \\ f_{n-1}(a) \end{bmatrix}$. Now because $\mathrm{rk}[L^T - aI] = n - 1$, it follows that there can only be one independent eigenvector for a, and thus there is exactly one Jordan block per eigenvalue, as expected, for a non-derogatory matrix.

In the simplest case $f(x) = \prod_{i=1}^{n}(x - \lambda_i)$ has n *distinct* roots and $L^T(f)$ has n distinct evalues, and as such is diagonalizable via its evector matrix $V = [X_n(\lambda_1), \ldots, X_n(\lambda_n)]$. Needless to say this is the celebrated Vandermonde matrix

$$V = \begin{bmatrix} 1 & 1 & & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & & & \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}.$$

We may conclude that if $f(x)$ has distinct roots then

$$V^{-1} L^T(f) V = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \Lambda$$

Likewise $L(f)$ is diagonalized by the matrix basis change GV, i.e. $L(GV) = (GV)D$, where $(GV)_{ij} = f_i(\lambda_j)$, $i, j = 0, 1, \ldots, n-1$.

We next note that, because $\frac{f(x)-f(y)}{x-y} = X_n' G_f Y_n$, if α and β are two distinct roots of $f(x)$ then the quotient vanishes and so $X_n'(\alpha) G_f Y_n(\beta) = 0$. Also letting x approach y, or by summing directly, we see that $X_n'(x) G_f Y_n(x) = f'(x)$. This means that $V^T G_f V = \mathrm{diag}(f'(\lambda_1), ., f')\lambda_n)) = D$ or $V^{-1} = DV^T G$, which can then be used to establish that

$$\begin{aligned} V^T B(f,g) V &= V^T g(L_f) G_f V = V^T G_f g(L_f^T) V \\ &= (V^T G V)\Lambda = D\Lambda. \end{aligned}$$

The companion matrix $\Omega = L(x^n - 1) = [\mathbf{e}_2, \ldots, \mathbf{e}_n, \mathbf{e}_1]$ is called the basic *circulant*, and any polynomial $p(\Omega)$ in Ω is a circulant matrix.

For example if $p(x) = p_0 + p_1 x + \cdots + p_{n-1}x^{n-1}$ then

$$p(\Omega) = \begin{bmatrix} p_0 & p_{n-1} & & p_2 & p_1 \\ p_1 & p_0 & p_{n-1} & \cdots & p_2 \\ p_2 & p_1 & p_0 & & \\ \vdots & & & & p_{n-1} \\ p_{n-1} & & \cdots & p_1 & p_0 \end{bmatrix}.$$

The matrix Ω is one of the most important matrices in all of applied mathematics. Indeed, since $\Delta_\Omega(\lambda) = \lambda^n - 1$, its eigenvalues are the n distinct n-th roots of unity, $\sigma = \{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$, where $\omega = exp(\frac{2\pi i}{n})$.

Consequently it also has n independent eigenvectors $\mathbf{v}_n(\omega^k)$, $k = 0, 1, \ldots, n-1$ (called phasers), and as such Ω can be diagonalized via

$$\Omega^T V = VD, \quad \text{and} \quad \Omega V = VD^{-1},$$

where $D = \operatorname{diag}(1, \omega, \ldots, \omega^{n-1})$ and V is the *Vandermonde* matrix

$$V = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & & \omega^{2(n-1)} \\ \vdots & & & & \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \ldots & \omega^{(n-1)(n-1)} \end{bmatrix} = V^T.$$

Now V is not only symmetric but its columns $\mathbf{v}(\omega^k)$, are also pairwise *orthogonal*! Indeed,

$$\mathbf{v}(\omega^k)^* \mathbf{v}(\omega^r) \quad = \quad \sum_{s=0}^{n-1} \omega^{s(r-k)} = \begin{cases} n & if \quad r = k \\ 0 & if \quad r \neq k. \end{cases}$$

Consequently we may normalize the eigenvectors and use the *unitary* matrix $W = \frac{1}{\sqrt{n}} V$ for which $W^{-1} = \bar{W}^T = \bar{W}$.

The matrix multiplication

$$\mathbf{y} = W\mathbf{x}$$

is referred to as the **Discrete Fourier Transform**. It is of cardinal importance in the theory of *filtering.* We shall now examine the case of repeated roots of $f(\lambda)$.

Like Janus, differentiation is a "two-faced" personality. On the one hand it is used to compute tangents and tangent planes, and as such is all important in optimization, while on the other hand it also serves as the **ultimate counting machine**. This makes it indispensable in combinatorics and in fact anywhere where polynomials are used. Recall that the term λ^k is after all just a place holder, and that its coefficient can be "counted" by differentiating k times. As such we have two counting tools, matrix multiplication and differentiation and our main trick is is to convert *differential* identities into *matrix* identities.

Suppose If $f(\lambda) = \prod_{i=1}^{s} (\lambda - \lambda_i)^{m_i} = (\lambda - \lambda_i)^{m_i} \phi_i(x)$. We shall now **differentiate** the companion shift in column form to solve this problem. Consider

$$L_f^T X_n(x) = X_n x - f(x)\mathbf{e}_n$$

and differentiate both sides k times. Using the product rule gives

$$L_f^T X_n^{(k)} = x X_n^{(k)} + k X_n^{(k-1)} - f^{(k)}(x)\mathbf{e}_n.$$

Dividing by $k!$ and setting $M_k = X_n/k!$, yields $L^T M_k = xM_k + M_{k-1} - (f^{(k)}/k!)\mathbf{e}_n$, which is a step down recurrence relation. Stacking r of these columns in $W_{r,n}(x) = [M_0, ., M_{r-1}]$ shows that

$$L^T W_{r,n}(x) = W_{r,n}(x) J_r(x) - \mathbf{e}_n F_r(x)^T,$$

where $F_r(x)^T = [f(x), \frac{f'(x)}{1!}, \ldots, \frac{f^{(r-1)}(x)}{(r-1)!}]$. If we substitute λ_i for x and take $r = m_i$, then $F(\lambda_i) = 0$ leaving $L^T W_{m_i,n}(\lambda_i) = W_{m_i,n}(\lambda_i) J_{m_i}(\lambda_i)$, where

$$W_{m,n+1}^T(\lambda) = \begin{bmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{m-1} & \binom{n}{n}\lambda^n \\ 0 & 1 & 2\lambda & \cdots & (m-1)\lambda^{m-2} & \binom{n}{n-1}\lambda^{n-1} \\ & & & & & \vdots \\ 0 & 0 & 1 & & & \\ 0 & 0 & 0 & \cdots & 1 & \binom{n}{n-m+1}\lambda^{n-m+1} \end{bmatrix}$$

is an $m \times (n+1)$ confluent Vandermonde block. Stacking these blocks for each of the distinct eigenvalues we arrive at

$$L^T[W_{\lambda_1}, \ldots, W_{\lambda_s}] = \\ [W_{\lambda_1}, \ldots, W_{\lambda_s}] \operatorname{diag}(J_{m_1}(\lambda_1), \ldots, J_{m_s}(\lambda_s)),$$

which is the desired Jordan form. Several points should now be noted:

(i) $W_{m,}(\alpha)$ precisely equals the chain matrix $K_n[\mathbf{e}_1, J_m^T(\alpha)]$.

(ii) it relates the derivatives to the coefficients of a polynomial in the stacked form

$$\begin{bmatrix} f(\lambda) \\ f'(\lambda)/1! \\ \vdots \\ \frac{f^{(m-1)}(\lambda)}{(m-1)!} \end{bmatrix} = W_{m,n}(\lambda) \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{bmatrix}.$$

(iii) The matrix $W = [W_{\lambda_1}, \ldots, W_{\lambda_s}]$ is the generalized Vandermonde matrix. It is also known as the Caratheodory matrix, which appears in the study of moment problems.

(iv) W also equals the Wronskian Matrix of the set of functions $\{t^j e^{\lambda_k t}\}$, $k = 1, \ldots, s$ and $j = 0, \ldots, m_k - 1$ and appears in the study of differential equations.

(v) If $[f_0, \ldots, f_n]W = 0$ then $f^{(j)}(\lambda_k) = 0$ for $k = 1, \ldots, s$ and $j = 0, \ldots, m_k$, with $m_1 + \cdots + m_s = n+1$. But then $\pi(x) = \prod_{i=1}^{s} (\lambda - \lambda_i)^{m_i} | f(x)$, in which $\partial(\pi) = n+1$ while $\partial(f) = n$. Thus forcing $f(x) = 0$, ensuring that W is non-singular. As a check we can compute

(vi) $\det(W) = \prod_{1 \le j < i \le s} (\lambda_i - \lambda_j)^{m_i m_j}$.

(vii) The generalized Vandermonde matrix also appears naturally in Hermite interpolation where one aims to find a polynomial $f(x)$ with prescribed derivatives at prescribed points.

Lastly, recall that $B(f,g) = g(L_f)G_f$, and suppose that $W^{-1} L_f^T W = J$ is the Jordan form of L. Then

$$\begin{aligned} W^T B(f,g) W &= W^T g(L_f) G_f W = W^T G_f [g(L_f^T) W] \\ &= \big(W^T G_f W\big) g(J). \end{aligned}$$

Now W^TGW is made up of blocks $W_\alpha^T G_f W_\beta$, where α and β are eigenvalues of L. When $\alpha \neq \beta$ we see from the difference quotient that this vanishes. On the other hand, when $\alpha = \beta$, we shall need more care. First recall that the adjoint polynomial satisfy $F_n' = X_n' G$, and hence that ${F_n'}^{(k)} = {X_n'}^{(k)} G$. Now since the rows of W_α^T are the derivatives of X_N at α we obtain

$$W_\alpha^T G_f = \begin{bmatrix} \frac{X_n'(\alpha)}{1!} \\ \vdots \\ \frac{{X_n'}^{(k-1)}(\alpha)}{(k-1)!} \end{bmatrix} G_f = \begin{bmatrix} \frac{F_n'(\alpha)}{1!} \\ \vdots \\ \frac{{F_n'}^{(k-1)}(\alpha)}{(k-1)} \end{bmatrix},$$

which is the weighted Wronskian of the adjoint polynomials at α. It is now clear that $(W_\alpha^T G_f W_\alpha)_{pq} = \frac{{F_n'}^{(p)} {X_n}^{(q)}}{p!q!}$. To compute this we first differentiate the adjoin shift equation (1.4) which gives

$$x{F_n'}^{k)} + kx{F_n'}^{k-1)} - {F_n'}^{k)} L^T = f^{(k)}(x)\mathbf{e}_1^T.$$

Now post mutiply by X_n, and substitute the column companion shift (0.1). This gives

$$k{F_n'}^{(k-1)}(x) X_n(x) = f^{(k)}(x),$$

where we used the fact that ${F_n'}^{(k)} \mathbf{e}_n = 0$. It now follows by induction that

$${F_n'}^{(k)} {X_n}^{(r)} = \frac{f^{(k+r+1)}(x) r!k!}{(r+k+1)!},$$

and thus the matrix $W_\alpha^T G_f W_\alpha$ can now be identified as $F\phi(J_{m_i}(\alpha))$.

10 The group Inverse of a Companion Matrix

We have seen that the inverse, if any, of a companion matrix L(f), again has companion structure. When L is singular we require in some settings the group inverse $L^\#$ (over a ring R with 1), which satisfies

$$LXL = L, \; XLX = X \text{ and } LX = XL.$$

It exists iff p_0 is regular and $w = p_0 - (1 - p_0 p_0^-) p_1$ is invertible, in which case it has the form $L^\# = [\mathbf{x}, \mathbf{y}, B]$ where $B = \begin{bmatrix} \mathbf{0}^T \\ I_{n-2} \\ \mathbf{0}^T \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} \mathbf{u} \\ x \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} \mathbf{v} \\ y \end{bmatrix}$, $\mathbf{v} = \mathbf{e}_1 + \hat{\mathbf{f}} y$ and $\mathbf{u} = \begin{bmatrix} v_2 \\ \vdots \\ v_{n-1} \\ y \end{bmatrix} + \hat{\mathbf{f}} x$ and $\hat{\mathbf{f}} = [p_1, \ldots, p_{n-1}]^T$.

The parameters x and y can be expressed in terms of w^{-1}, p_0, p_1 and p_2. Its structure is again sparse, but is closer to that of a perturbed companion matrix. The expression for the Drazin inverse however, is still unknown.

11 Conclusions

We have seen that the companion shift equation is central to many of the applications involving L. It is in combination with other shift conditions that cancellation can occur and the best results materialize. There are numerous generalizations of a companion a matrix, such as the comrade and congenial matrices which use other bases besides the powers of x. Many of the chain relations generalize to the block case and provide a inexhaustible supply of challenging problems.

Bibliography

[1] S. Barnett, *Division of Generalized Polynomials using the Comrade matrix*, Lin ALg. App. 60 (1984), pp. 159-75.

[2] —, *A matrix circle in Linear Control*, 12 Bull. Ins. maths. Appl. 9 (1976), pp. 173-177.

[3] L. Brand, *The companion matrix and its properties*, Math. Monthly, vol 71 (1964), pp. 629-634.

[4] G. Chen and Z. Yang, *Bezoutian Representation via Vandermonde Matrices*, Lin ALg. App, 186 (1993), pp. 37-44.

[5] R. E. Hartwig, *Resultants, Companion Matrices and Jacobson Chains*, Lin ALg App 94 (1987), pp. 1-34.

[6] ——, *AX - XB = C, Resultants and generalized inverses*, SIAM J. App. Math. 28 (1975), pp. 154-183.

[7] G. Heinig, *Matrix Representations of Bezoutians*, Lin ALg. App, 223/224 (1995), pp. 337-54.

[8] U. Helmke and P.A. Fuhrmannn, *Bezoutians*, Lin ALg. App, 122 (1989), pp. 1039-1097.

[9] R. Puystjens and R.E.Hartwig, *The Group Inverse of a Companion Matrix*, Lin.Mul.Lin. Alg. 43 (1997), pp.137-150.

[10] N. J. Young, *Linear Fractional Transform of Companion Matrices*, Glasgow Math. J. 20 (1979), pp. 129-132.