Companion matrices and their relations to Toeplitz and Hankel matrices

Yousong Luo* and Robin Hill

Abstract: In this paper we describe some properties of companion matrices and demonstrate some special patterns that arise when a Toeplitz or a Hankel matrix is multiplied by a related companion matrix. We present a necessary and sufficient condition, generalizing known results, for a matrix to be the transforming matrix for a similarity between a pair of companion matrices. A special case of our main result shows that a Toeplitz or a Hankel matrix can be extended using associated companion matrices, preserving the Toeplitz or Hankel structure respectively.

Keywords: Companion matrix, Toeplitz matrix, Hankel matrix, Bezoutian

AMS subject classifications: 15A21, 15A24, 15A99

1 Introduction and notation

Companion matrices occur not only in matrix analysis but also in many scientific fields. For example, a companion matrix $C_1$ naturally arises as the system matrix when a dynamic system is represented in state space form [9, 12]. When a basis of the state vector space is changed a new system matrix appears and, quite often, the new system matrix is also a companion matrix $C_2$ in different form and the similarity relation between the new and the old system matrices is realized by a nonsingular basis changing matrix $P$, i.e. $PC_1P^{-1} = C_2$ or

$$PC_1 = C_2P.$$  

Here we will call the basis changing matrix $P$ the transformation matrix. In the literature on dynamic systems the transformation matrices are constructed in order to have a better understanding of the state vectors from a certain aspect, and hence they are verified case by case for the similarity between the pair of companion matrices. A question arises naturally: If $C_1$ and $C_2$ are a given pair of similar companion matrices, what is in common among those transformation matrices $P$ satisfying (1)? The most obvious approach is to solve equation (1) directly for all solutions $P$. Although (1) is not hard to solve it is still not clear, after finding the solutions, what is in common among these $P$'s.

In this note we give a necessary and sufficient condition for $P$ to satisfy (1) for certain pairs of similar companion matrices $C_1$ and $C_2$. This condition is not deep but it is clear and constructive. We believe that it is worthwhile to add this little piece of information into the existing rich theory in the literature of matrix analysis. As a byproduct of this condition we have found a pattern in the extension of Toeplitz or Hankel matrices by powers of associated companion matrices, preserving the Toeplitz or Hankel structure respectively. This relation has been actually used implicitly in [2] and [4] where the determinant of an extended Toeplitz matrix is computed via powers of determinants of certain companion matrices. Finally, as an application, we give

*Corresponding Author: Yousong Luo: School of Mathematical and Geospatial Sciences, RMIT University, GPO Box 2476V, Melbourne, Vic. 3001, AUSTRALIA, E-mail: yousong.luo@rmit.edu.au
Robin Hill: Department of Electrical and Electronic Engineering, University of Melbourne, Vic. 3010, AUSTRALIA, E-mail: robinh1@unimelb.edu.au

© 2015 Yousong Luo and Robin Hill, licensee De Gruyter Open.
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.
an example linking the state of a dynamic system after any number of iterations to the initial state together
with knowledge of the input up to the current time, using just one matrix being the extension of the system
matrix by powers of companion matrices.

Given vectors \( \mathbf{u} = (u_1, \ldots, u_{n+1})^T \) and \( \mathbf{v} = (v_1, \ldots, v_{n+1})^T \in \mathbb{R}^{n+1} \) we define the polynomials
\[
    u(\lambda) = u_1 + u_2 \lambda + \cdots + u_{n+1} \lambda^{n-1} + u_{n+1} \lambda^n
\]
and
\[
    v(\lambda) = v_1 + v_2 \lambda + \cdots + v_{n+1} \lambda^{n-1} + v_{n+1} \lambda^n.
\]
We assume always that \( u_1, u_{n+1}, v_1 \) and \( v_{n+1} \) are nonzero, and that \( u(\lambda) \) and \( v(\lambda) \) are co-prime. The “top”,
“bottom”, “left” and “right” companion matrices of the polynomial \( u(\lambda) \) (or the vector \( \mathbf{u} \)) are defined as
\[
    C_t(\mathbf{u}) := \begin{bmatrix}
        -\frac{u_n}{u_{n+1}} & \cdots & -\frac{u_1}{u_{n+1}} \\
        \frac{1}{u_{n+1}} & \cdots & 0
    \end{bmatrix}, \quad
    C_b(\mathbf{u}) := \begin{bmatrix}
        0 & I_{n-1} & -u_2 \\
        \frac{u_{n+1}}{u_1} & \cdots & -u_2
    \end{bmatrix}
\]
and
\[
    C_l(\mathbf{u}) := \begin{bmatrix}
        \frac{u_2}{u_1} & I_{n-1} \\
        \vdots & \ddots & \vdots \\
        -\frac{u_{n+1}}{u_1} & 0 & -u_2
    \end{bmatrix}
\]
When their dependence on \( \mathbf{u} \) is clear from context we will simply write \( C_t, C_b, C_l \) and \( C_r \). The companion
matrices of \( v(\lambda) \) are defined in the same way. Under our assumptions on \( \mathbf{u} \) and \( \mathbf{v} \), all the companion matrices
defined above are nonsingular.

Let \( f \) be the flipping matrix
\[
    f = \begin{bmatrix}
        0 & 1 \\
        \vdots & \ddots \\
        1 & 0
    \end{bmatrix}
\]
For a vector \( \mathbf{u} \) we denote by \( \mathbf{u}^f \) the vector \( f \mathbf{u} \), and the corresponding polynomial \( u^f(\lambda) \) is defined by \( u^f(\lambda) = u_{n+1} + u_n \lambda + \cdots + u_2 \lambda^{n-1} + u_1 \lambda^n \). For a matrix \( A \) we denote by \( A^f \) the flipping of \( A \) about its secondary diagonal, so \( A^f = fA^Tf \). Hankel matrices are symmetric in the usual sense but Toeplitz matrices \( A \) are persymmetric,
that is, symmetric about their secondary diagonal
\[
    A^f = A.
\]
We also define the companion matrices of \( \mathbf{u}^f \) and denote them by \( C_t(\mathbf{u}^f), C_b(\mathbf{u}^f), C_l(\mathbf{u}^f) \) and \( C_r(\mathbf{u}^f) \). When their dependence on \( \mathbf{u}^f \) is clear from context we will simply write these matrices as \( \overline{C}_t, \overline{C}_b, \overline{C}_l \) and \( \overline{C}_r \).

Define the following triangular Toeplitz matrices using the components of \( \mathbf{u} \) and \( \mathbf{v} \):
\[
    U_+ := \begin{bmatrix}
        u_1 & 0 & \cdots & 0 \\
        u_2 & u_1 & \ddots & \vdots \\
        \vdots & \ddots & \ddots & 0 \\
        u_n & \cdots & u_2 & u_1
    \end{bmatrix}, \quad
    U_- := \begin{bmatrix}
        u_{n+1} & u_n & \cdots & u_2 \\
        0 & u_{n+1} & \ddots & \vdots \\
        \vdots & \ddots & \ddots & u_n \\
        0 & \cdots & 0 & u_{n+1}
    \end{bmatrix}
\]
Analagously we define \( V_+ \) and \( V_- \) in terms of the components of \( \mathbf{v} \).

The Toeplitz Bezoutian \( B_T := \text{Bez}_T(\mathbf{u}, \mathbf{v}) = (b_{ij})_{i,j=1}^n \) and Hankel Bezoutian \( B_H := \text{Bez}_H(\mathbf{u}, \mathbf{v}) = (c_{ij})_{i,j=1}^n \)
of the vectors \( \mathbf{u}, \mathbf{v} \) (or the polynomials \( u(\lambda), v(\lambda) \)) are the \( n \times n \) matrices with the generating polynomials
\[
    \sum_{i,j=1}^n b_{ij} \lambda^{i-1} \mu^{j-1} = \frac{u(\lambda) v(\mu) - u^f(\mu) v(\lambda)}{1 - \mu \lambda}
\]
and
\[
    \sum_{i,j=1}^n c_{ij} \lambda^{i-1} \mu^{j-1} = \frac{u(\lambda) v(\mu) - u(\mu) v(\lambda)}{\lambda - \mu}
\]
respectively. The Gohberg-Semencul formulae [5, 6] imply that the Toeplitz Bezoutian matrix generated by \( u \) and \( v \) is
\[
\mathbf{B}_T = U_* V_+ - V_+ U_* = V_+ U_+ - U_+ V_+ ,
\] (5)
and the Hankel Bezoutian matrix generated by \( u \) and \( v \) is
\[
\mathbf{B}_H = V_* J U_+ - U_+ J V_* = U_+ J V_+ - V_+ J U_+ .
\] (6)
It is known [10] that if \( u(\lambda) \) and \( v(\lambda) \) are co-prime then \( \mathbf{B}_T \) and \( \mathbf{B}_H \) are both nonsingular and that \( \mathbf{B}_T^{-1} \) is Toeplitz and \( \mathbf{B}_H^{-1} \) is Hankel.

\section{Similarity of companion matrices}

We first list some obvious relations among the companion matrices defined in Section 1.

\textbf{Properties:}

1. Inversion:
\[
\mathbf{C}_t = \mathbf{C}_b^{-1}, \quad \mathbf{C}_l = \mathbf{C}_r^{-1}, \quad \mathbf{C}_t = \mathbf{C}_b^{-1}, \quad \mathbf{C}_l = \mathbf{C}_r^{-1} .
\] (7)

2. Flipping:
\[
\mathbf{C}_t^T = \mathbf{C}_u, \quad \mathbf{C}_b^T = \mathbf{C}_l, \quad \mathbf{C}_t^T = \mathbf{C}_r, \quad \mathbf{C}_b^T = \mathbf{C}_l .
\] (8)

3. Transposition:
\[
\mathbf{C}_t^T = \mathbf{C}_l, \quad \mathbf{C}_b^T = \mathbf{C}_r, \quad \mathbf{C}_r^T = \mathbf{C}_b, \quad \mathbf{C}_l^T = \mathbf{C}_l .
\] (9)

All these properties can be easily verified. Property 1 can also be found in [3].

\subsection{Similarity}

In this section we derive a necessary and sufficient condition for a nonsingular matrix \( \mathbf{P} \) to satisfy equation (1).

For convenience we define a simple operation on square Toeplitz or Hankel matrices. For an invertible Toeplitz matrix
\[
\mathbf{T} = \begin{bmatrix}
    a_0 & a_{-1} & \cdots & a_{1-n} \\
    a_1 & a_0 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & a_{-1} \\
    a_{n-1} & \cdots & a_1 & a_0 \\
\end{bmatrix},
\] (10)
the \( (n - 1) \times (n + 1) \) Toeplitz matrix \( \partial \mathbf{T} \), introduced in [7], is obtained by adding one column to the right preserving the Toeplitz structure and then deleting the first row:
\[
\partial \mathbf{T} := \begin{bmatrix}
    a_1 & a_0 & a_{-1} & \cdots & a_{1-n} \\
    \vdots & \ddots & \ddots & \ddots & \vdots \\
    a_{n-1} & \cdots & a_1 & a_0 & a_{-1} \\
\end{bmatrix} .
\] (11)

Similarly, for an invertible Hankel matrix \( \mathbf{H} = \mathbf{T} J \), the \( (n - 1) \times (n + 1) \) Hankel matrix \( \partial \mathbf{H} \) is obtained by adding one column to the right preserving the Hankel structure and then deleting the last row:
\[
\mathbf{H} = \begin{bmatrix}
    a_{1-n} & \cdots & a_{-1} & a_0 \\
    \vdots & \ddots & \ddots & \vdots \\
    a_{-1} & \ddots & \ddots & \vdots \\
    a_0 & a_1 & \cdots & a_{n-1} \\
\end{bmatrix} , \quad \partial \mathbf{H} := \begin{bmatrix}
    a_{1-n} & \cdots & a_{-1} & a_0 & a_1 \\
    \vdots & \ddots & \ddots & \ddots & \vdots \\
    \vdots & \ddots & \ddots & \ddots & \vdots \\
    a_{-1} & a_0 & a_1 & \cdots & a_{n-1} \\
\end{bmatrix} .
\] (12)

In the following, given a matrix \( \mathbf{A} \), we will use \( \mathbf{A}_{[i,j,k,l]} \) to denote the sub-matrix of \( \mathbf{A} \) formed by selecting all rows from the \( i \)th row to the \( j \)th row and all columns from the \( k \)th column to the \( l \)th column.
Theorem 2.1. The following three statements are equivalent.
1. $T$ is an invertible Toeplitz matrix and $u \in \text{Ker}(\partial T)$.
2. $T$ is an invertible matrix satisfying
   \[ C_r T = T C_r. \]  
   (13)
3. $T$ is an invertible matrix satisfying
   \[ C_r T = T C_1. \]  
   (14)

In such a case $C_r T$, $T C_r$, $C_r T$ and $T C_1$ are all Toeplitz matrices.

Proof. We first prove statement 2 implies statement 1. Let $A = C_r T$, $B = T C_r$ and $A = B$. Due to the structure of $C_r$ and $C_1$, we can easily see that $A_{[2,n,1:n]} = T_{[1:n-1,1:n]}$ and $B_{[1:n,1:n-1]} = T_{[1:n,2:n]}$ which implies that $A_{[2,n,1:n-1]} = T_{[1:n-1,1:n-1]}$ and $B_{[2,n,1:n-1]} = T_{[2,n,2:n]}$. By the assumption $A = B$ we obtain $T_{[1:n-1,1:n-1]} = T_{[2,n,2:n]}$ and hence $T$ is Toeplitz. Now we can write $T$ in the form of (10). Then we have
\[
\partial T u = T_{[2,n,1:n]} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + u_{n+1} \begin{bmatrix} a_{1-n} \\ \vdots \\ a_{-1} \end{bmatrix} = -u_{n+1} B_{[2,n,n:n]} + u_{n+1} \begin{bmatrix} a_{1-n} \\ \vdots \\ a_{-1} \end{bmatrix}.
\]

Now $B_{[2,n,n:n]} = A_{[2,n,n:n]} = T_{[1:n-1,1:n]} = \begin{bmatrix} a_{1-n} & \cdots & a_{-1} \end{bmatrix}^T$. Putting this into the equation above gives
\[
\partial T u = 0, \quad \text{so } u \in \text{Ker}(\partial T).
\]

Now we prove statement 1 implies statement 2. It is easy to see that
\[
T C_r = \begin{bmatrix} T_{[1:n,2:n]} & \beta \end{bmatrix}
\]
where $\beta$ is a column given by
\[
\beta = -\frac{1}{u_{n+1}} T_{[2,n,1:n]} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = -\frac{1}{u_{n+1}} \begin{bmatrix} T_{[1:n,1:n]} \\ T_{[2,n,1:n]} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}.
\]

Since $u \in \text{Ker}(\partial T)$ we have
\[
T_{[2,n,1:n]} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + u_{n+1} \begin{bmatrix} a_{1-n} \\ \vdots \\ a_{-1} \end{bmatrix} = 0.
\]
This implies that
\[
\beta = \begin{bmatrix} \mu_{-1} \\ a_{1-n} \\ \vdots \\ a_{-1} \end{bmatrix}
\]
where $\mu_{-1} = -\frac{1}{u_{n+1}} \begin{bmatrix} a_0 & a_{-1} & \cdots & a_{1-n} & 0 \end{bmatrix} u$, that is, $\beta^T J$ is the first row of $T C_r$. From this, by equation (15) and the assumption that $T$ is Toeplitz, $T C_r = \begin{bmatrix} T_{[1:n,2:n]} & \beta \end{bmatrix}$ is Toeplitz and hence $T C_r = (T C_r)^\dagger$. On the other hand, since $T$ is Toeplitz, by equation (8) we have
\[
(T C_r)^\dagger = J (T C_r)^T J = (J (C_r)^T J) (J^T J) = (C_r)^T T^\dagger = C_r T.
\]
As a consequence we have that both $T C_r$ and $C_r T$ are Toeplitz and
\[
T C_r = C_r T.
\]

The proof for the equivalence of statement 1 and statement 3 is similar.
To prove statement 3 implies statement 1 we begin by assuming \( A = T C_l , B = C_b T \) and \( A = B \) as before. The structure of \( C_l \) and \( C_b \) implies that \( A_{[1:n−1,2:n]} = T_{[1:n−1,1:n−1]} \) and \( B_{[1:n−1,2:n]} = T_{[2:n,2:n]} \). Then from \( A = B \) we obtain \( T_{[1:n−1,1:n−1]} = T_{[2:n,2:n]} \) which means that \( T \) is Toeplitz. By equating the first columns of \( A \) and \( B \) we get

\[
-\frac{1}{u_1} T \begin{bmatrix} u_2 \\ \vdots \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_{n-1} \\ \mu_1 \end{bmatrix}
\]

where \( \mu_1 = -\frac{1}{u_1} \begin{bmatrix} 0 & a_{n-1} & \cdots & a_1 & a_0 \end{bmatrix} \). Deleting the last row on both sides we obtain \((\partial T) u = 0\).

Finally we prove statement 1 implies statement 3. The structure of \( C_l \) gives

\[
TC_l = \begin{bmatrix} \gamma & T_{[1:n,1:n−1]} \end{bmatrix}
\]

where \( \gamma \) is a column given by

\[
\gamma = \frac{1}{u_1} T \begin{bmatrix} u_2 \\ \vdots \\ u_{n+1} \end{bmatrix} = \frac{1}{u_1} T_{[1:n−1,1:n]} \begin{bmatrix} T_{[1:n−1,1:n]} \\ T_{[n:n,1:n]} \end{bmatrix} \begin{bmatrix} u_2 \\ \vdots \\ u_{n+1} \end{bmatrix}.
\]

Since \( u \in \text{Ker}(\partial T) \) we have

\[
u_1 \begin{bmatrix} a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} + T_{[1:n−1,1:n]} \begin{bmatrix} u_2 \\ \vdots \\ u_{n+1} \end{bmatrix} = 0.
\]

This implies that

\[
\gamma = \begin{bmatrix} a_1 \\ \vdots \\ a_{n-1} \\ \mu_1 \end{bmatrix}
\]

where \( \mu_1 = -\frac{1}{u_1} \begin{bmatrix} 0 & a_{n-1} & \cdots & a_1 & a_0 \end{bmatrix} \), that is, \( \gamma^T J \) is the last row of \( TC_l \). From this, by equation (16) and the assumption that \( T \) is Toeplitz, \( TC_l = \begin{bmatrix} \gamma & T_{[1:n,1:n−1]} \end{bmatrix} \) is Toeplitz and hence \( TC_l = (TC_l)^J \). On the other hand, since \( T \) is Toeplitz, by equation (8) we have

\[
\]

As a consequence we have

\[
TC_l = C_b T.
\]

\break

**Corollary 2.2.** The following three statements are equivalent.
1. \( H \) is an invertible Hankel matrix and \( u^J \in \text{Ker}(\partial H) \).
2. \( H \) is an invertible matrix satisfying

\[
C_l H = H C_l.
\]

3. \( H \) is an invertible matrix satisfying

\[
C_b H = H C_r.
\]

In such a case \( C_l H, H C_l, C_b H \) and \( H C_r \) are all Hankel matrices.
Proof. Let $T = HJ$. Then $T$ is a non-singular Toeplitz matrix. Obviously $u^T ∈ \text{Ker}(∂H)$ is equivalent to $u ∈ \text{Ker}(∂T)$. Now we prove that $TC_r = C_lT$ is equivalent to $C_lH = HC_l$. Since $H$ is Hankel we then have 

$$T^l = JT^lJ = (TJ)^TJ = H^TJ = HJ.$$  

By taking the flipping operation on both sides of $TC_r = C_lT$ we obtain $C_l^TJ^l = T^lC_l$ which is, by (8) and Equation (19), $C_lH = HC_l^TJ^l$. Then multiplying both sides by $J$ gives 

$$C_lH = HHC_l.$$

By now we can see that Theorem 2.1 and Corollary 2.2 are basically the same thing.

In the remainder of this section we give two examples. They provide insight into the use of similarity transformations in dynamic systems.

**Example 1.** Our condition is constructive, so we can easily build a transformation matrix in a particular pattern that we want without having to solve the matrix equation $C_bT = TC_l$. It is clear from Theorem 2.1 that there are an infinite number of nonsingular matrices $T$ which satisfy this equation. In order to arrive at the one used in the systems and control literature it is necessary to constrain $T$ to be lower triangular. Suppose we want to find a lower triangular $T$ satisfying $C_bT = TC_l$ for the case $u = [1, 2, 3, 4]^T$. Then 

$$C_b(u) := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & -3 & -2 \end{bmatrix} \quad \text{and} \quad C_l(u) := \begin{bmatrix} -2 & 1 & 0 \\ -3 & 0 & 1 \\ -4 & 0 & 0 \end{bmatrix}.$$  

The lower triangular $T$ is automatically Toeplitz and $∂T$ takes the form 

$$\begin{bmatrix} b & a & 0 & 0 \\ c & b & a & 0 \end{bmatrix}.$$  

From the condition of Theorem 2.1 $u ∈ \text{Ker}(∂T)$ we can simply put (since $[a, b, c] = [t, −2t, t]$ for any $t ∈ \mathbb{R}$) $a = 1$, $b = −2$ and $c = 1$ which gives 

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

We can see immediately that this $T$ is actually $U^{-1}_s$ and it confirms the following situation in the field of dynamic system. In [9] the canonical observer form of a dynamic system has, in our notation, system matrix $A_o = C_l(u)$, and the canonical observability form of the same dynamic system has system matrix $A_{ob} = C_b(u)$. The transformation matrix determined by the dynamics in [9] is $U_s$ such that $A_oU_s = U_sA_{ob}$. In terms of our Theorem 2.1, this is exactly $TC_l = C_bT$ with $T = U_s^{-1}$.

**Example 2.** In this example we apply our condition to another example from [9] where the transformation matrix is constructed in a quite complicated way from the dynamics. Consider a linear dynamic system represented in the state space form 

$$\dot{x}(t) = Ax + Bw, \quad y = Cx,$$

where $x(t)$ is the state vector, $A$ is the system matrix, $B$ is the input column matrix, $C$ is the output row matrix, $w$ is a scalar input and $y$ is the scalar output. The state space representation is in canonical controller form if the system matrix $A_c = C_l(u)$ with $u = (a_n, \ldots, a_1, 1)^T$, the input matrix $B_c = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T$ and the output matrix $C_c = \begin{bmatrix} b_1 & \ldots & b_n \end{bmatrix}$. The system representation is in canonical observer form if $A_o = C_l(u^T) = \bar{C}_l(u)$, $B_o = C_l^T$ and the output matrix is $C_o = B_o^T$. The controllability matrices $C(A_c, B_c)$, $C(A_o, B_o)$ and the observability matrices $O(C_c, A_c)$, $O(C_o, A_o)$ are then constructed in the usual way. Under the assumption that the system is both controllable and observable it is shown that the matrix 

$$Q = O^{-1}(C_o, A_o)O(C_c, A_c) = C(A_o, B_o)C^{-1}(A_c, B_c)$$
will transform \( A_0 \) into \( A_c \) by way of \( Q^{-1}A_0Q = A_c \), that is
\[
QC_i = C_iQ. \tag{20}
\]

It is also shown in [9] that \( Q = -JBT \) where \( B_T \) is the Toeplitz Bezoutian \( \text{Bez}_T(u, v) \) where \( v = (b_0, \ldots, b_1, 0)^T \).

Our condition provides insight into this complicated situation. For this we need some results from Corollary 2.3, 2.10, Theorem 4.2 and 4.5 of [6]. We merely summarize the relevant information in the following

\textbf{Theorem 2.3.} A necessary and sufficient condition for two non-zero Toeplitz Bezoutian matrices \( B_T(a, b) \) and \( B_T(a_1, b_1) \) to coincide is
\[
\begin{bmatrix}
  a_1 & b_1 \\
  a & b
\end{bmatrix} \varphi
\]
for some matrix \( \varphi \) with \( \det \varphi = 1 \).

If \( T \) is an invertible Toeplitz matrix and \( \{a, b\} \) is a basis for the kernel of \( \partial T \), then \( B_T(a, b) \) is just a scalar multiple of \( T^{-1} \).

Let \( T = B_T^{-1} \) where \( B_T = \text{Bez}_T(u, v) \). It is well known that \( T \) is Toeplitz, (see [6]), and hence
\[
H = -TJ = -B_T^{-1}J = (-JBT)^{-1} = Q^{-1}
\]
is Hankel. Putting \( T = B_T^{-1} \) in Theorem 2.3 we see that if \( \{a, b\} \) is a basis for the kernel of \( \partial(B_T^{-1}) \) then \( B_T(a, b) = \lambda T^{-1} = \lambda B_T(u, v) \) for some nonzero constant \( \lambda \). Up to a scaling of the vector \( a \) we may assume that \( B_T(a, b) = B_T(u, v) \). By Theorem 2.3 again we have
\[
\begin{bmatrix}
  u & v \\
  a & b
\end{bmatrix} \varphi
\]
for some invertible matrix \( \varphi \). This means that both \( u \) and \( v \) are in the kernel of \( \partial T \). It follows that \( u^t \) is in the kernel of \( \partial H \). Finally by equation (17) of Corollary 2.2 we have
\[
C_i(u)H = H\overline{C}_i(u) \tag{21}
\]
which is (20).

\section{3 Extension using companion matrices}

\subsection{3.1 Extension}

A Toeplitz (or Hankel) matrix can be extended to any size in a Toeplitz (or Hankel) way, by adding more diagonal bands to the existing bands. Theoretical aspects of such an extension, such as the minimum rank of the extension, have been studied in the literature (see [1] and the references therein). What we are concerned with here is a specific way of extending the matrix by multiplication by some associated companion matrices. We hope this extension might have more applications than the ones we will demonstrate at the end of this paper. We will use the similarity relations among companion matrices that have been developed earlier.

Assume \( A \) is an \( n \times n \) matrix. The role that \( C_i \) plays in the product \( C_iA \) is to keep the first \( n - 1 \) rows of \( A \) as the last \( n - 1 \) rows of \( C_iA \), and to add one new row on the top. The new row added is a linear combination of rows of \( A \). Similarly, the first \( n - 1 \) rows of \( C_kA \) are the last \( n - 1 \) rows of \( A \) and the last row of \( C_kA \) is a linear combination of rows of \( A \). This enables us to extend the matrix \( A \) in the upward and downward directions as follows. Starting from \( A \), for integers \( k \geq 1 \) we define \( \mathcal{T}[A : k, l] \) to be the \( (n + k - l) \times n \) matrix
\[
\mathcal{T}[A : k, l] = \begin{bmatrix}
\gamma_{k-1} \\
\vdots \\
\gamma_1 \\
C_iA
\end{bmatrix} \tag{22}
\]
where $\gamma_j (i=1,2,\ldots,k-l)\text{ is the first row of } C_{i+l}^j A$.

In similar fashion the effect of post-multiplying a matrix by $C_i$ or $C_l$ can be considered. We can extend a matrix in the right and left directions by adding the last column of $AC_i$ to the right or adding the first column of $AC_l$ to the left. Starting from $\mathcal{T}[A : k, l; s, t]$, for integers $s \geq t$ we define

$$\mathcal{T}[A : k, l; s, t] = \begin{bmatrix} \mathcal{T}[A : k, l] C_l^t & \cdots & \beta_1 \cdots & \beta_{s-t} \end{bmatrix},$$

where $\beta_j (i=1,2,\ldots,s-t)\text{ is the last column of } \mathcal{T}[A : k, l] C_l^{s-t}$. We call this a Toeplitz extension because we will prove that, under certain conditions, such an extension preserves the Toeplitz structure if the starting matrix $A$ is Toeplitz. We call $A$ a generator in such an extension. To generate the same matrix $\mathcal{T}[A : k, l; s, t]$ we can use any $n \times n$ matrix of the form $C_i^t AC_l^t$, where $i$ and $j$ are integers. It is easy to see that

$$\mathcal{T}[A : k, l; s, t] = \mathcal{T}[C_i^t AC_l^t : k-i, l-i; s-j, t-j].$$

If we use $C_i^t$ and $C_l^t$ instead of $C_i$ and $C_l$ in the above extension, we will obtain a different extended matrix $\mathcal{X}[A : k, l; s, t]$. We call this a Hankel extension because it preserves the Hankel structure under certain conditions.

Here are two examples:

$$\mathcal{X}[I : n, -n; n, -n] = \begin{bmatrix} C^n_t A^n_t & C^n_t A^n_t & C^n_t A^n_t \\
C^n_t A^n_t & I & C^n_t \\
C^n_t A^n_t & C^n_t & C^n_t \\
\end{bmatrix},$$

$$\mathcal{X}[A : n, -n; -n, -2n] = \begin{bmatrix} C^n_t AC^n_t & C^n_t AC^n_t & C^n_t AC^n_t \\
AC^n_t & AC^n_t & AC^n_t \\
C^n_t AC^n_t & C^n_t AC^n_t & C^n_t AC^n_t \\
\end{bmatrix}.$$

We notice that, for any square matrix $A$,

$$\mathcal{T}[A : k, l; s, t] = \mathcal{T}[I : k, l; 0, 0] A \mathcal{T}[I : 0, 0; s, t].$$

An obvious property of these extensions is given in the following Proposition.

**Proposition 3.1.** Suppose $A$ is invertible. Then the rank of $\mathcal{T}[A : k, l; s, t]$ is $n$. If $r = s - t > 0$ then $\{e_1, \ldots, e_r\}$ is a basis for the kernel of $\mathcal{T}[A : k, l; s, t]$, where $e_i$ is the $i$th column of the $(n + r) \times r$ Toeplitz matrix whose first column is $[u_1 \cdots u_{n+1} 0 \cdots 0]^T$ and last column is $[0 \cdots 0 u_1 \cdots u_{n+1}]^T$. In particular, $\mathcal{T}[A : k, k; s, s - 1] e_i = 0$ for all integers $k$ and $s$.

**Proof.** All the rows of $\mathcal{T}[A : k, l; s, t]$ are linear combinations of the rows of $\mathcal{T}[A : 0, 0; s, t]$ which is a rank $n$ matrix. Thus $\mathcal{T}[A : k, l; s, t]$ is of rank $n$. It also follows that the kernel of $\mathcal{T}[A : k, l; s, t]$ is the same as the kernel of $\mathcal{T}[A : 0, 0; s, t]$. The latter is an $n \times (n + r)$ matrix so its kernel has dimension $r$. It is clear that the set $\{e_1, \ldots, e_r\}$ is linearly independent. So the only thing we need to verify is $\mathcal{T}[A : 0, 0; s, t] e_i = 0$ for $i = 1, \ldots, r$. Due to the structure of $e_i$

$$\mathcal{T}[A : 0, 0; s, t] e_i = \begin{bmatrix} AC_{s}^{i+1} & e_i \end{bmatrix} u$$

where $u$ is the last column of $AC_{s}^{i+1} = AC_{s}^{i+1} A_r$. Therefore $e = AC_{s}^{i+1} b$ where $b$ is the last column of $C_r$. A direct verification yields $\begin{bmatrix} I & b \end{bmatrix} u = 0$. As a consequence

$$\begin{bmatrix} AC_{s}^{i+1} & e \end{bmatrix} u = \begin{bmatrix} AC_{s}^{i+1} & AC_{r}^{i+1} \end{bmatrix} \begin{bmatrix} I \\ b \end{bmatrix} u = 0.$$

**Corollary 3.2.** Suppose $A$ is invertible. Then the rank of $\mathcal{X}[A : k, l; s, t]$ is $n$. If $r = s - t > 0$ then $\{e'_1, \ldots, e'_r\}$ is a basis for the kernel of $\mathcal{X}[A : k, l; s, t]$, where $e'_i$ is the $i$th column of the $(n + r) \times r$ Hankel matrix whose first column is $[0 \cdots 0 u_{n+1} \cdots u_1]^T$ and last column is $[u_{n+1} \cdots u_1 0 \cdots 0]^T$. 


Proof. Use \( C_i \) and \( e_i^j \) instead of \( C_r \) and \( e_j \) in the proof of Proposition 3.1.

The following Lemma is a preparation for the proof of our main Theorem 3.5.

**Lemma 3.3.** Let \( T \) be a Toeplitz matrix and \( u = (u_1, \ldots, u_{n+1})^T \) be a vector such that \( u_1, u_{n+1} \neq 0 \). If \( u \) belongs to the kernel of \( \partial T \) then \( u \) also belongs to the kernels of \( \partial(C_i T), \partial(C_b T), \partial(T C_i) \) and \( \partial(T C_j) \).

**Proof.** We only prove the case \( \partial(C_i T) \); the proof for the case \( \partial(C_b T) \) is similar. The other two cases are covered by Theorem 2.1. Let

\[
T = \begin{bmatrix}
a_0 & \cdots & a_{1-n} \\
\vdots & \ddots & \vdots \\
a_{n-1} & \cdots & a_0
\end{bmatrix}
\]

then, By Theorem 2.1, \( C_i T \) is Toeplitz and hence

\[
C_i T = \begin{bmatrix}
a_{-1} & \cdots & a_{1-n} & \mu_{-1} \\
\end{bmatrix}_{T[1,n-1,n]}
\]

where

\[
\mu_{-1} = -\frac{1}{u_{n+1}}[u_n, \ldots, u_1][a_{1-n}, \ldots, a_0]^T.
\]

It follows that

\[
\partial(C_i T) = \begin{bmatrix}
a_0 & \cdots & a_{1-n} & \mu_{-1} \\
\end{bmatrix}_{S_1}
\]

where \( S_1 \) is the sub-matrix of \( \partial T \) consisting of the first \( n-2 \) rows of \( \partial T \). From the definition of \( \mu_{-1} \) we have

\[
\begin{bmatrix}
a_0 & a_{-1} & \cdots & a_{1-n} & \mu_{-1}
\end{bmatrix}u = 0
\]

and hence

\[
\partial(C_i T)u = 0.
\]

Putting \( H = TJ \), we have immediately

**Corollary 3.4.** Let \( H \) be a Hankel matrix and \( u = (u_1, \ldots, u_{n+1})^T \) be a vector such that \( u_1, u_{n+1} \neq 0 \). If \( u \) belongs to the kernel of \( \partial H \) then \( u \) belongs to the kernels of \( \partial(C_i H) \) and \( \partial(C_b H) \).

The more interesting features of the extensions are now presented.

**Theorem 3.5.** Let \( T \) be an invertible Toeplitz matrix and \( u = (u_1, \ldots, u_{n+1})^T \) be a vector such that \( u_1, u_{n+1} \neq 0 \). If \( u \) belongs to the kernel of \( \partial T \), then the matrix \( T[T : k, l; s, t] \) is Toeplitz.

**Proof.** Because any \( n \times n \) block in \( T[T : k, l; s, t] \) is in the form of \( C_i^j T C_{j}^i \) for some integers \( i \) and \( j \), we only need to prove that all such blocks are Toeplitz. We use the same argument in all the four directions of extension and only demonstrate the detailed argument in two directions, namely the upward and rightward directions.

We prove the Toeplitz extension in upward direction first. Without loss of generality we assume \( j = 0 \) and we prove that \( C_i^j T \) is Toeplitz by induction on \( i > 0 \). Theorem 2.1 has already covered the case \( i = 1 \). Assume now all matrices \( C_i^j T, s = 0, 1, \ldots, i \) are Toeplitz. Since \( u \) belongs to the kernel of \( \partial T \) by the hypothesis, Lemma 3.3 guarantees that \( u \) belongs to the kernel of \( \partial(C_i T) \). Then applying Lemma 3.3 successively gives that \( u \) belongs to the kernel of \( \partial(C_{j}^i T) \) for each \( s = 0, 1, \ldots, i \). Then by Theorem 2.1 we conclude that \( C_i^{j+1} T \) is Toeplitz.

Now we prove the extension in rightward direction. From the proof above, \( C_i^j T \) is Toeplitz \( u \) belongs to the kernel of \( \partial(C_i T) \) for all \( i \geq 0 \). Assume now all matrices \( C_i^j T C_{s}^j, s = 0, 1, \ldots, j \) are Toeplitz. Since \( u \) belongs to the kernel of \( \partial(C_i T) \) by the proof above, Lemma 3.3 guarantees that \( u \) belongs to the kernel of \( \partial(C_{j}^i T C_{s}^j) \).
Applying Lemma 3.3 successively gives that $u$ belongs to the kernel of $\partial(C^j_i TC^j_i)$ for each $s = 0, 1, \ldots, j$. Finally by Theorem 2.1 we conclude that $C^j_i TC^j_i$ is Toeplitz. Therefore we have completed the induction for the upward and rightward directions.

In the downward and leftward directions we have negative indices $i < 0$ and $j < 0$ in $C^j_i TC^j_i$. By repeating the same induction argument for the negative indices from $i$ to $i - 1$ and from $j$ to $j - 1$, we have the Toeplitz extension in the downward and leftward directions.

**Corollary 3.6.** Let $H$ be an invertible Hankel matrix and $u = (u_1, \ldots, u_{n+1})^T$ be a vector such that $u_1, u_{n+1} \neq 0$. If $u^i$ belongs to the kernel of $\partial H$, then the matrix $\mathcal{H}([H : k, l, s, t])$ is Hankel.

**Proof.** Define $T = H/|H|$ then $T$ satisfies the conditions of Theorem 3.5 and hence this Corollary follows. □

As a summary of this section we an example of Toeplitz extension.

**Example** $\mathcal{T}[U^{-1}_s : k, l; s, t]$. $U^{-1}_s$ is lower triangular and we denote its elements in the first column by $s_1, \ldots, s_n$. Then $\partial U^{-1}_s$ has the form

$$
\partial U^{-1}_s := \begin{bmatrix} s_2 & s_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ s_n & \cdots & s_2 & s_1 & 0 \end{bmatrix}
$$

and, obviously, $u$ is in the kernel of $\partial U^{-1}_s$. Therefore, $\mathcal{T}[U^{-1}_s : k, l; s, t]$ is Toeplitz. If $m = \max\{k, s\} > 0$ and $h = \max\{l, t\} > 0$, it can be shown that such an extension has the following features: (a) there is a middle band consisting of $n - 1$ diagonal lines of zeros; (b) the elements below the middle band of zeros are the elements (displayed in the same order) of the inverse of the $(n + h) \times (n + h)$ lower triangular Toeplitz matrix whose first column is $[u_1, \ldots, u_{n+1}, 0, \ldots, 0]^T$; (c) the elements above the middle band of zeros are the elements (displayed in the same order) of the inverse of the $m \times m$ upper triangular Toeplitz matrix whose first row is the truncation of the first $m$ elements of $[-u_{n+1}, \ldots, -u_1, 0, \ldots]$. Due to the Toeplitz structure of such an extension we only need to prove these features for the cases when $k, l, s$ and $t$ are integer multiples of $n$ and the general case is just an arbitrary truncation of these special cases. We only demonstrate the proof in the case of $k = s = t = n$ and $l = 0$. We write

$$
\mathcal{T}[U^{-1}_s : n; 0; n] = \begin{bmatrix} S_1 & R_1 & R_2 \\ S_2 & S_1 & R_1 \end{bmatrix} = \begin{bmatrix} S_1 & 0 & 0 \\ S_2 & S_1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & R_1 & R_2 \\ 0 & 0 & R_1 \end{bmatrix}.
$$

It is easy to see that

$$
\begin{bmatrix} S_1 & 0 \\ S_2 & S_1 \end{bmatrix} = \begin{bmatrix} U_+ & 0 \\ U_- & U_+ \end{bmatrix}^{-1}
$$

because $S_1 = U^{-1}_s$ and, by Proposition 3.1, $S_2 U_+ + S_1 U_- = 0$. We now show that

$$
\begin{bmatrix} R_1 & R_2 \\ 0 & R_1 \end{bmatrix} = - \begin{bmatrix} U_- & U_+ \\ 0 & U_- \end{bmatrix}^{-1}.
$$

To see this we apply Proposition 3.1 to $\mathcal{T}[U^{-1}_s : n; 0; n]$:

$$
0 = \begin{bmatrix} S_1 & 0 & 0 \\ S_2 & S_1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & R_1 & R_2 \\ 0 & 0 & R_1 \end{bmatrix} + \begin{bmatrix} U_+ & 0 \\ U_- & U_+ \end{bmatrix} - \begin{bmatrix} U_- & U_+ \\ 0 & U_- \end{bmatrix}.
$$

The first term on the right hand side is equal to $I$ and hence the second term on the right hand side is equal to $-I$. 

3.2 Applications

Here we give an example of application of the above extension. This is a problem studied in [8] and we briefly describe it as follows. For an arbitrary sequence \(y = (y_k)_{k=1}^\infty\), its \(\lambda\)-transform (generating function) is defined to be \(\hat{y}(\lambda) := \sum_{k=1}^\infty y_k \lambda^{-k}\). Consider a linear, time-invariant, causal, discrete-time dynamic system with transfer function description \(\hat{y}(\lambda) = -(v(\lambda)/u(\lambda))\hat{w}(\lambda)\), where \(w = (w_k)_{k=1}^\infty\) is the input, \(y = (y_k)_{k=1}^\infty\) is the output, and the numerator and the denominator of the transfer function \(-v(\lambda)/u(\lambda)\) satisfy all the assumptions stated in Section 1, that is, \(u(\lambda) = u_1 + u_2 \lambda + \cdots + u_{n+1} \lambda^n\) and \(v(\lambda) = v_1 + v_2 \lambda + \cdots + v_{n+1} \lambda^n\) with nonzero \(u_1, u_{n+1}, v_1, v_{n+1}\) as well as that \(u(\lambda)\) and \(v(\lambda)\) are co-prime. We will represent this system in state space form by introducing state vectors first and then write down the rule of evolution of state vectors in terms of the initial state and the input.

When both \(w\) and \(y\) are sequences in \(l_1\), the analytic functions \(\hat{w}(\lambda)\) and \(\hat{y}(\lambda)\) are related by

\[ u(\lambda)\hat{y}(\lambda) = -v(\lambda)\hat{w}(\lambda). \]

Equating like powers of \(\lambda\) gives

\[ Uy + Vw = 0 \]

where \(w\) and \(y\) are column vectors and

\[
U := \begin{bmatrix}
U_+ & 0 & 0 \\
U_- & U_+ & \ddots \\
0 & \ddots & \ddots
\end{bmatrix}, \quad V := \begin{bmatrix}
V_+ & 0 & 0 \\
V_- & V_+ & \ddots \\
0 & \ddots & \ddots
\end{bmatrix}.
\]

It can be shown using functional analysis arguments that such \(w\) and \(y\) have the form

\[(w, y) = (-Ub, Vb) \text{ for some } b \in l_\infty. \tag{24}\]

Writing out this equation in detail yields the difference equations

\[ w_k = -u_{n+1}b_k - u_nb_{k+1} - \cdots - u_2b_{k+n-1} - u_1b_{k+n} \tag{25} \]

and

\[ y_k = v_{n+1}b_k + v_nb_{k+1} + \cdots + v_2b_{k+n-1} + v_1b_{k+n}. \tag{26} \]

Now we can introduce naturally the \(n\)-dimensional state vector at the time \(k\) as the truncation \(b_{[k:n+1]}\) of \(b \in l_\infty\) and denote it by

\[ x(k) = (b_k, b_{k+1}, \ldots, b_{k+n-1})^T. \]

Then we can put (25) and (26) in the state space form

\[ x(k + 1) = Ax(k) + Bw_k, \quad y_k = Cx(k) - \frac{v_1}{u_1}w_k \tag{27} \]

where the system matrix is \(A = Cb\), the input matrix \(B\) is

\[
B = \begin{bmatrix}
0 \\
\vdots \\
0 \\
1/u_1
\end{bmatrix},
\]

and the output matrix \(C\) is

\[
C = \frac{1}{u_1} \begin{bmatrix}
u_1v_{n+1} - u_1u_{n+1} & \cdots & u_1v_2 - u_1u_2
\end{bmatrix}.
\]
Note that the output matrix $C$ is actually the first row of $B_T(u, v)$ divided by $u_1$. The general truncation $b_{[1:n+p]} = [b_1, \ldots, b_{n+p}]^T$ for $p > 0$ is then given by

$$b_{[1:n+p]} = \mathcal{T}[I: 0, -p; 0, 0]x(0) - \frac{1}{u_1} \begin{bmatrix} O_{n+p} \\ F_p \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_p \end{bmatrix},$$

(28)

where

$$F_p = \begin{bmatrix} 1 & 0 \\ s_1 & 1 \\ \vdots & \ddots & \ddots \\ s_{p-1} & \cdots & s_1 & 1 \end{bmatrix},$$

and $s_i$ is the element at the last column and last row of $(C_i)^i$. It can be shown that $(1/u_1)F_p$ is the inverse of the $p \times p$ lower triangular nonsingular truncation of $U$, but we skip the proof here.

Now we change the basis of the state space by using the transformation matrix $B_T(u, v)$, that is, we introduce $\dot{x}(k) = B_Tx(k)$. Then, by (21), the state space representation of the system is transformed into another canonical form

$$\dot{x}(k + 1) = A'\dot{x}(k) + B'w_k, \quad y_k = C'\dot{x}(k) - \frac{v_1}{u_1}w_k$$

(29)

where the system matrix $A'$ = $C_i$, and the input and output matrices become

$$B' = -\frac{1}{u_1} \text{ Last column of } B_T \quad \text{and} \quad C' = \frac{1}{u_1} \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}.$$

Then the state vector at the time $q > 0$ can be expressed directly in terms of the initial state $\dot{x}(0)$ and the input data:

$$\dot{x}(q) = (A')^q\dot{x}(0) + \mathcal{T}[I: 0, 0; 0, 1 - q]E_q \begin{bmatrix} w_1 \\ \vdots \\ w_q \end{bmatrix},$$

(30)

where $E_q$ is the $(n + q - 1) \times q$ band matrix

$$\begin{bmatrix} B_1 & 0 \\ & \ddots \\ & \phantom{1} & B_1 \end{bmatrix}.$$

A combination of (28) and (30) gives the following mixed case. Evolving the state vectors under the first basis until a given time $q$, then changing into the second basis and evolving further to the time $q + p$, we have

$$b_{[q+1:n+q+p]} = \mathcal{T}[I: 0, -p; 0, 0]B_T^{-1}\dot{x}(q) - \frac{1}{u_1} \begin{bmatrix} O_{n+p} \\ F_p \end{bmatrix} \begin{bmatrix} w_{q+1} \\ \vdots \\ w_{q+p} \end{bmatrix}$$

$$= \mathcal{T}[B_T^{-1}: 0, -p; n - q, -q]x(0) + \mathcal{T}[I: 0, -p; 0, 1 - q]E_q \begin{bmatrix} w_1 \\ \vdots \\ w_q \end{bmatrix} - \frac{1}{u_1} \begin{bmatrix} O_{n+p} \\ F_p \end{bmatrix} \begin{bmatrix} w_{q+1} \\ \vdots \\ w_{q+p} \end{bmatrix}$$

(31)

for all positive integers $p$ and $q$. As we can see that using our extension we can link the state variable at any time directly to the initial state by one equation without having to express it in an iterative way. All extensions involve powers of companion matrices. The formula derived in [11] can be used to calculate entries of an integer power of a companion matrix directly.
References


