Systems & Control Letters 4 (1984) 243–251
North-Holland

July 1984

Reduced order observers: A new algorithm and proof

Paul Van DOOREN

Philips Research Laboratory, Av. Van Becelaere 2, B-1170 Brussels, Belgium

Received 22 March 1984

This paper gives a new recursive algorithm to construct a reduced order observer (with prescribed spectrum) of a given observable system. The method is based on the staircase form and implicitly gives a new proof for the existence of such reduced order observers.

Keywords: Numerical methods, Reduced order observer, Staircase form, Schur form, Sylvester equation.

1. Introduction

The reduced order observer problem can be formulated as follows. Let the input, state and output dimension of the system (here λ denotes the differential operator in the continuous-time case and the shift operator in the discrete-time case)

$$\lambda x = Ax + Bu, \qquad y = Cx \tag{1}$$

be m, n and p, respectively, and let r be the rank of C, then, under the assumption that the multivariable state-space system $\{A, B, C\}$ is observable, one can construct a reduced order observer [8], [9] of the form

$$\lambda z = Fz + Dy + Pu, \tag{2}$$

using the input u and output y of the system (1). The system (2) has state dimension $n-r$, and the estimated state

$$\hat{x} = [G \vdots H]\begin{bmatrix} z \\ \hdashline y \end{bmatrix} \tag{3}$$

converges asymptotically to the state x if F, D, P, G and H are appropriately chosen. The reduced order observer configuration is illustrated in Figure 1.

In the sequel we will assume, without loss of generality, that C has full now rank, i.e. $r = p$ (see also Remark 2(c)). A constructive method to find the matrices F, D, P, G and H is then given by the following (modified) theorem, due to Luenberger [9].

Theorem 1. *Let T be an $(n-r) \times n$ matrix such that*

$$TA - FT = DC, \qquad \begin{bmatrix} T \\ \hdashline C \end{bmatrix} \textit{invertible}. \tag{4a;b}$$

Then putting

$$P = TB, \qquad [G \vdots H] = \begin{bmatrix} T \\ \hdashline C \end{bmatrix}^{-1}, \tag{5a;b}$$

we have that

$$\hat{x}(t) - x(t) = Ge^{Ft}[z(0) - Tx(0)]. \tag{6}$$

A solution T for (4) *does exist for any choice of spectrum for F. Convergence of the reduced order observer is thus obtained when choosing F in* (4) *to be stable.* □

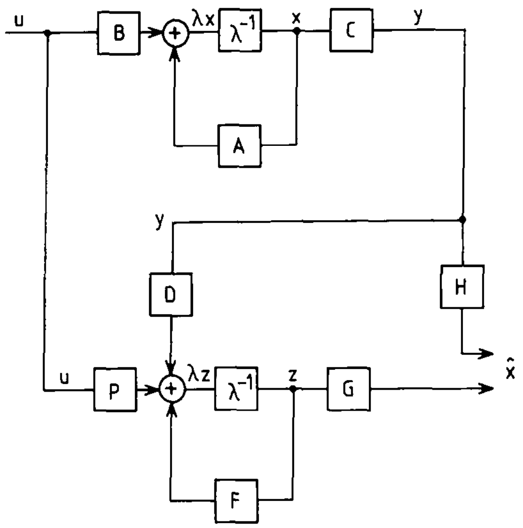

Fig. 1.

In [9], an algorithm is also provided for solving T satisfying (4), when the spectrum of F has already been chosen. The method, though, is based on the decomposition of the multivariable system (1) into a set of single input systems. This canonical decomposition should be avoided if possible, for reasons of numerical stability, and methods for solving T directly from (4) should be used instead. Such an alternative method is suggested in [8] but is then restricted to the case where A and F have no common eigenvalues and is based on the Jordan canonical form of A, which again ought to be avoided for reasons of numerical stability. A better approach for solving for T is suggested in [3], where A and F are supposed to be in some type of condensed form that can be obtained via the use of unitary similarity transformations. In this paper we give a new method based on the staircase form of the system (1) and which is strongly inspired from the Hessenberg–Schur method described in [3].

2. Problem (re)formulation

The method described in this paper is based on the use of unitary state-space transformations of the systems (1) and (2), chosen because of their invariance property with respect to certain norms:

$$\|UAV\| = \|A\| \quad \text{for } U, V \text{ unitary, i.e. } U^*U = UU^* = I,\ V^*V = VV^* = I,$$

where $\|\cdot\|$ stands for both the spectral and Frobenius norms [12], and * denotes the (conjugate) transpose of a (complex) matrix. To start with, the (A, C)-pair is transformed via a unitary state-space transformation U to 'staircase form' (see e.g. [1], [7], [11], [13], [14], [15], [17]):

$$A_U = U^*AU = \begin{array}{c} \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & \cdots & A_{1,k} \\ A_{2,1} & A_{2,2} & & & A_{2,k} \\ & \ddots & \ddots & & \vdots \\ & 0 & \ddots & \ddots & \vdots \\ & & & A_{k,k-1} & A_{k,k} \end{bmatrix} \end{array} \begin{array}{l} \}r_1 \\ \}r_2 \\ \\ \\ \}r_k \end{array}, \tag{7a}$$

$$C_U = CU = \Big[\underbrace{\quad\quad}_{r_1}\ \underbrace{\quad\quad}_{r_2}\ \ 0\ \ \underbrace{\quad\quad}_{r_{k-1}}\ \underbrace{\ C_k\ }_{r_k}\Big]\}\, m = r_{k+1}. \tag{7b}$$

Here C_k and the $A_{i,i-1}$ off-diagonal blocks have full column rank r_i by construction. Moreover, these blocks can be chosen to be in triangular form [10]:

$$\underbrace{\left.\begin{bmatrix} X & x & \dots & \dots & x \\ 0 & X & & & x \\ \vdots & & \ddots & & \vdots \\ & & & X & x \\ 0 & & \dots & 0 & X \\ 0 & & \dots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & & \dots & 0 & 0 \end{bmatrix}\right\} r_{i+1}}_{r_i} \tag{8}$$

where the X's are nonzero elements and the x's are arbitrary.

Equation (4) is now equivalent to

$$TUU^*AU - FTU = DCU, \qquad \begin{bmatrix} TU \\ \hline CU \end{bmatrix} \text{ invertible}, \tag{9a;b}$$

or, with $T_U = TU$, to

$$T_U A_U - FT_U = DC_U, \qquad \begin{bmatrix} T_U \\ \hline C_U \end{bmatrix} \text{ invertible}. \tag{10a;b}$$

Similarly we can define, for an arbitrary unitary transformation V, a new equivalent equation

$$T_{VU} A_U - F_V T_{VU} = D_V C_U, \qquad \begin{bmatrix} T_{VU} \\ \hline C_U \end{bmatrix} \text{ invertible}, \tag{11a;b}$$

where

$$T_{VU} = V^* T_U, \qquad F_V = V^* F V, \qquad D_V = V^* D. \tag{12}$$

Therefore we can restrict ourselves to finding a solution T_{VU} for F_V in (lower) Schur form, since for an arbitrary F we can always choose V such that F_V is in that form. Note that by an appropriate choice of the state-space transformations U and V of the respective systems (1) and (2), we have thus deduced an equivalent set of equations which is more 'condensed' [16]. This will lead to the elegant recursive solution described in the next section. But it is equally important to note here that by our specific choice of U and V, we obtain an equivalent numerical problem. By this we mean that the sensitivities of the solutions T of (4) and T_{VU} of (11) are the same, the reason for this being the unitarity of U and V (see also [3], [16]).

Remark 1. If the system (1) is real, a real observer can be found as well [9]. All the above matrices can then be chosen real, with the only restriction that F_V can only be reduced to its real (lower) Schur form, i.e. a quasi triangular form with a 2×2 bump on the diagonal to each pair of complex conjugate eigenvalues of F_V.

Since F is to be constructed and only its spectrum is specified, one can immediately require it to be in its (lower) Schur form (i.e. $V = I$), and thus put $T_{VU} = T_U$, $F_V = F$ and $D_V = D$. This will be assumed in the sequel.

3. A recursive algorithm

In this section we show that in the coordinate system of (10), there exists a solution T_U of the form

$$T_U = \left.\begin{bmatrix} 1 & x & \cdots & & & & & \cdots & x \\ 0 & 1 & x & & & & & & x \\ \vdots & & \ddots & \ddots & & & & & \vdots \\ & & & \ddots & \ddots & & & & \\ \vdots & & & & 1 & x & & & \vdots \\ 0 & \cdots & & \cdots & 0 & 1 & x & \cdots & x \end{bmatrix}\right\} n-r. \qquad \text{with the last } r \text{ columns marked } \underbrace{}_{r} \tag{13}$$

Thereby, A_U, C_U and the spectrum of F are assumed to be given, while F itself, D and T_U are to be constructed. The proof of the existence of such a solution T_U is given by construction.

For simplicity, we first assume all matrices to be complex since then the Schur form of F is strictly triangular. From the proportion of the staircase form (7), (8) and from (13) it immediately follows that the compound matrix (10b) is invertible since it is upper triangular with nonzero diagonal elements. That a solution T_U of the form (13) satisfies also (10a) is now shown recursively. We therefore denote the diagonal elements of F (i.e. its spectrum) by f_{ii} and the i-th rows of F, T_U and D respectively as (here $'$ is used to denote a row vector):

$$[\underbrace{f_i'}_{i-1} \quad \underbrace{f_{ii}}_{1} \quad \underbrace{0 \ \ldots \ 0}_{n-r-i}], \qquad [\underbrace{0 \ \ldots \ 0}_{i-1} \quad \underbrace{1}_{1} \quad \underbrace{t_i'}_{n-i}], \qquad \underbrace{d_i'}_{r}. \tag{14}$$

Using this, the first row of (10a) becomes

$$\begin{bmatrix} 1 & t_1' \end{bmatrix} A_U - f_{11} \begin{bmatrix} 1 & t_1' \end{bmatrix} = d_1' C_U \tag{15}$$

and analogously for the i-th row of (10a),

$$\begin{bmatrix} 0 & \ldots & 0 & 1 & t_i' \end{bmatrix} A_U - \begin{bmatrix} f_i' & f_{ii} & 0 & \ldots & 0 \end{bmatrix} T_U = d_i' C_U. \tag{16}$$

These can now be rewritten as

$$\begin{bmatrix} t_1' & d_1' \end{bmatrix} \begin{bmatrix} Af_1 \\ \hline -C_U \end{bmatrix} = -af_1' \tag{17}$$

and

$$[\underbrace{f_i'}_{i-1} \quad \underbrace{t_i'}_{n-i} \quad \underbrace{d_i'}_{r}] \begin{bmatrix} -T_i \\ \hline Af_i \\ \hline -C_u \end{bmatrix} = -af_i', \tag{18}$$

where Af_i and af_i' denote respectively the bottom $n-i$ rows and the i-th row of the matrix $(A_U - f_{ii}I)$, and T_i denotes the top $i-1$ rows of T_U:

$$(A_U - f_{ii}I_n) = \begin{bmatrix} X \\ af_i' \\ Af_i \end{bmatrix} \begin{matrix} \}i-1 \\ \}1 \\ \}n-i \end{matrix}, \qquad T_U = \begin{bmatrix} T_i \\ \\ X \end{bmatrix} \begin{matrix} \}i-1 \\ \\ \}n-r-i+1 \end{matrix} \tag{19a,b}$$

Notice that (17) and (18) denote a system of equations of the type

$$x_i' M_i(x_1 \quad \ldots \quad x_{i-1}) = af_i' \tag{20}$$

in the unknown $x_i' = [f_i' \quad t_i' \quad d_i']$, whereby M_i depends on the previous vectors x_i'. One can thus solve recursively for the rows of F, T_U and D using (17) and then (18) for $i = 2, \ldots, r$, provided all of the M_i matrices have a left inverse.

We now show that the required invertibility property of the M_i matrices is always fulfilled because of the assumed staircase form of the (A_U, C_U)-pair. Indeed, a typical configuration of an M_i matrix would be (at step $i = 4$) for an (A_U, C_U)-pair with stairs $r_1 = 2$, $r_2 = 2$, $r_3 = 3$ and $r_4 = 3$,

$$\left[\begin{array}{c} A_u \\ \hline -C_u \end{array}\right] = \left[\begin{array}{cccccccccc}
x & x & x & x & x & x & x & x & x & x \\
x & x & x & x & x & x & x & x & x & x \\
X & x & x & x & x & x & x & x & x & x \\
0 & X & x & x & x & x & x & x & x & x \\
0 & 0 & X & x & x & x & x & x & x & x \\
0 & 0 & 0 & X & x & x & x & x & x & x \\
0 & 0 & 0 & 0 & x & x & x & x & x & x \\
0 & 0 & 0 & 0 & X & x & x & x & x & x \\
0 & 0 & 0 & 0 & 0 & X & x & x & x & x \\
0 & 0 & 0 & 0 & 0 & 0 & X & x & x & x \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & X & x & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X
\end{array}\right],$$

$$\left[\begin{array}{c} -T_i \\ \hline Af_i \\ \hline -C_u \end{array}\right] = \left[\begin{array}{cccccccccc}
-1 & x & x & x & x & x & x & x & x & x \\
0 & -1 & x & x & x & x & x & x & x & x \\
0 & 0 & -1 & x & x & x & x & x & x & x \\
\hline
0 & 0 & X & x & x & x & x & x & x & x \\
0 & 0 & 0 & X & x & x & x & x & x & x \\
0 & 0 & 0 & 0 & x & x & x & x & x & x \\
0 & 0 & 0 & 0 & X & x & x & x & x & x \\
0 & 0 & 0 & 0 & 0 & X & x & x & x & x \\
0 & 0 & 0 & 0 & 0 & 0 & X & x & x & x \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & X & x & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X
\end{array}\right].$$

Both the above matrices are in a kind of column echelon form: each column has a leading non-zero element (from the bottom) with a different row index. For the left matrix this follows from the staircase form. For the right matrix (i.e. M_i) we remark that its bottom part is of the same form as that of the left matrix and that its top part is that of T_U, which is of the form (13). Each M_i matrix is thus in column echelon form and therefore left invertible. This then proves that for any choice of spectrum $\{f_{ii}, i = 1, \ldots, n-r\}$ of F, one can construct a solution T_U, F and D to (10) and thus to the reduced order observer problem. For this purpose we have not made any use of Luenberger's decomposition into single output systems as in [9], but rather we have used the properties of the staircase form. We remark, however, that the staircase form is in fact a first step toward such a decomposition since it gives the observability indices of the (A, C)-pair [14]. It is shown here that it is not needed to go beyond that first step. As a matter of fact, this ought to be discouraged since a further reduction to such a canonical decomposition requires numerically unstable transformations [13], [14] and then possibly leads to an uncontrolled build-up of rounding errors.

In the present approach one can also use the freedom of choice in T_U, F and D in order to solve the recursive systems (20) in minimum norm sense, which is a logical choice since it minimizes the norm of the matrix D and of the off diagonal part of F and T_U. The latter two will yield a better condition number for the eigenvalue problem of F [12], [18] and of the inversion of the compound matrix (10b), two numerical properties that are welcome in this problem.

In order to construct the minimum norm solution of the equations (17), (18), one performs the QR

decomposition of the matrices M_i [12], [18]:

$$M_i = Q_i R_i. \tag{21}$$

Using Householder transformations, this requires approximately rn^2 'flops' (1 flop = 1 addition + 1 multiplication) per M_i matrix, whereby one exploits the 'condensed' form of these matrices [16] (they have at most $r-1$ nonzero subdiagonals). For the solution of the triangularized system R_i and the back transformation with Q_i one needs approximately $\frac{1}{2}n^2$ and $2rn$ flops, respectively, per vector x_i', which is thus negligible with respect to the work required for the decomposition (21). Since there are $(n-r)$ such decompositions this amounts to (roughly):

$$(n-r)rn^2 \text{ flops} \quad \text{for solving for } F, T_U \text{ and } D. \tag{22}$$

For obtaining the staircase form (7) one needs approximately [16]

$$(3n+r)n^2 \text{ flops} \quad \text{including the construction of } U. \tag{23}$$

Finally, for constructing G, H and P as in (5) we use

$$P = T_U U^* B \quad \text{and} \quad [G \,\vdots\, H] = U \begin{bmatrix} T_U \\ \hline C_U \end{bmatrix}^{-1}, \tag{24}$$

which requires (using the special structure of T_U and C_U)

$$n^2 m + (n-r)nm \quad \text{and} \quad n^3/2 \quad \text{flops,} \tag{25}$$

respectively.

In the real case these operation counts vary slightly because of the occurrence of 2×2 bumps on the diagonal of F, corresponding to the pairs of complex conjugate eigenvalues of F. Each time such a bump is met in the rows of F, one has to replace the two corresponding systems (18) by a single system of doubled dimension. This is obtained as follows. Let $a + jb$ and $a - jb$ be the two complex conjugate eigenvalues of F. They can be realized by a 2×2 diagonal block:

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \tag{26}$$

Assume this block has row indices i and $i+1$ in F. Denote the rows i and $i+1$ of T_U and D as before and those of F as (note that f_{i+1}' is now one element shorter)

$$\begin{bmatrix} f_i' & a & -b & 0 & \dots & 0 \\ f_{i+1}' & b & a & 0 & \dots & 0 \end{bmatrix}. \tag{27}$$

Then, proceeding as before, one has to solve the matrix equation

$$\begin{bmatrix} 0 & \dots & 0 & 1 & t_i' \\ 0 & \dots & 0 & 1 & t_{i+1}' \end{bmatrix} A_U - \begin{bmatrix} f_i' & a & -b & 0 & \dots & 0 \\ f_{i+1}' & b & a & 0 & \dots & 0 \end{bmatrix} T_U = \begin{bmatrix} d_i' \\ d_{i+1}' \end{bmatrix} C_U, \tag{28}$$

or, equivalently,

$$\begin{bmatrix} 0 & \dots & 0 & 1 & t_i' \\ 0 & \dots & 0 & 1 & t_{i+1}' \end{bmatrix} A_U - \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} 0 & \dots & 0 & 1 & t_i' \\ 0 & \dots & 0 & 1 & t_{i+1}' \end{bmatrix} = \begin{bmatrix} f_i' \\ f_{i+1}' \end{bmatrix} T_i + \begin{bmatrix} d_i' \\ d_{i+1}' \end{bmatrix} C_U. \tag{29}$$

This is now again a linear system in the entries of f_i', t_i', d_i', f_{i+1}', t_{i+1}' and d_{i+1}'. Using the notation of Kronecker products, (29) can be rewritten as a linear system of the type (20):

$$xb_{i,i+1}' Mb_{i,i+1} = -\left(ab_i' + ab_{i+1}'\right) \tag{30}$$

with $xb'_{i,i+1}$ the vector that contains the (alternating) entries of the unknown vectors,

$$
\begin{aligned}
xb'_{i,i+1} = [\, & f_{i+1,1} \quad f_{i,1} \quad \cdots \quad f_{i+1,i-1} \quad f_{i,i-1} \\
& t_{i,i+1} \quad t_{i+1,i+2} \quad t_{i,i+2} \quad \cdots \quad t_{i+1,n} \quad t_{i,n} \quad d_{i+1,1} \quad d_{i,1} \quad \cdots \quad d_{i+1,r} \quad d_{i,r}],
\end{aligned}
\tag{31}
$$

and with

$$
Mb_{i,i+1} = \begin{bmatrix} -Tb_i \\ \hdashline Ab_i \\ \hdashline -Cb_U \end{bmatrix} \begin{matrix} \}2(i-1) \\ \}2(n-i)-1 \\ \}2r \end{matrix}
\tag{32}
$$

Here $Tb_i = T_i \otimes I_2$, $Cb_U = C_U \otimes I_2$ and ab'_i, ab'_{i+1} and Ab_i are the $2i$-th, the $(2i+1)$-th and the remaining $2(n-i)-1$ bottom rows, respectively, of the following Kronecker product:

$$
\begin{matrix} 2i-1\{ \\ 1\{ \\ 1\{ \\ 2(n-i)-1\{ \end{matrix}
\begin{bmatrix} X \\ \hdashline ab'_i \\ \hdashline ab'_{i+1} \\ \hdashline Ab_i \end{bmatrix} = (A_U - aI_n) \otimes I_2 - I_n \otimes \begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix}.
\tag{33}
$$

Because of the properties of the staircase form again, it is easily checked that the matrix $Mb_{i,i+1}$ is a $2(n+r)-3 \times 2n$ matrix with a column echelon structure similar to M_i. It is thus also left invertible and has now at most $2r-3$ nonzero subdiagonals. For the example used above, a complex eigenvalue pair at step $i=4$ would yield a matrix of the form

$$
Mb_{4,5} = \left|\begin{array}{cccccccccccccccccccc}
-1 & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 \\
0 & -1 & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x \\
0 & 0 & -1 & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 \\
0 & 0 & 0 & -1 & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & -1 & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x \\
\hdashline
0 & 0 & 0 & 0 & 0 & X & 0 & x & -b & x & 0 & x & 0 & x & 0 & x & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & b & x & 0 & x & 0 & x & 0 & x & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & -b & x & 0 & x & 0 & x & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & 0 & x & 0 & x & b & x & 0 & x & 0 & x & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & 0 & x & -b & x & 0 & x & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & 0 & x & b & x & 0 & x & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & -b & x & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & 0 & x & b & x & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & -b & x & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & 0 & x & b \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & -b & x \\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X
\end{array}\right|
$$

The number of flops needed to solve the system (33) in least squares sense is about $8rn^2$ since it has twice as many rows and columns as the comparable system (20). If all eigenvalues of F are chosen to be complex conjugate, one thus needs approximately:

$$4(n-4)rn^2 \text{ flops} \quad \text{for solving for } F,\ T_U \text{ and } D, \tag{34}$$

i.e. 4 times as much as would be the case for real eigenvalues. In the case of real F, it is thus recommended to choose real eigenvalues only.

Remark 2. (a) The above operation counts can be misleading in the sense that for complex eigenvalues of F, 4 times as much operations are needed when using real arithmetic (because of the 2×2 bumps) as when using complex arithmetic (where F has no 2×2 bumps anymore). This is of course not the case since a complex flop is 4 times as much work as a real flop.

(b) The most time-consuming step of the overall algorithm depends on the number of inputs and outputs: for small r and m (e.g. for SISO) the reduction to staircase from requires most of the work. For r larger than 3 the construction of T_U, D and F takes over most of the computational burden. The latter even becomes of the order of n^4 when r is comparable to n. This can be avoided by solving (20) and (30) directly instead of in least squares sense. This is indeed feasible without QR decomposition since the M and Mb matrices are in echelon form: a mere backsubstitution will yield 'a' solution for x' and xb' in $\frac{1}{2}n^2$ and $2n^2$ flops, respectively. For the full set of eigenvalues this thus becomes of the order of n^3. This of course may result in a more sensitive solution as was discussed earlier.

(c) The restriction $r = p$ (i.e. C has full row rank) does not affect the generality of our results, since a rank reduction of C can be obtained by the staircase form in case $r < p$: by allowing a unitary output transformation as well, C_k in (7b) can be put in the form (8). This displays the linear dependency of the columns of C and one uses then the above above algorithm on a transformed and reduced C matrix with r linearly independent columns.

4. Conclusion

In this paper we derived a new method and proof for the reduced order observer problem. The appeal of the present approach is that it does not require canonical forms or decompositions (as e.g. [4], [8], [9]) nor the computations of eigenvalues (as e.g. [8]): the method is fully recursive and is based on the staircase form of the system to be observed. Alternative methods, such as solving a Sylvester equation in T after having chosen F and D [8], [9] or solving a reduced feedback problem [2] (which could be done with [10]), do not exploit the degrees of freedom as is done here: the method indeed tries to find a solution that also minimizes in a certain sense the sensitivity of the solution. In regard to this, there is still more work to be done: the form of the constructed solution T_U is, although convenient, also a restriction on the problem. Optimization methods of the type used in [5] could be used to further improve the sensitivity of the solution for a general form of T_U. Finally, no control is possible here on the eigenstructure of F (i.e. its Jordan blocks in the case of repeated eigenvalues) since the off-diagonal elements of F are computed by the algorithm. That and other extensions of the reduced order observer problem are considered in [6] (but without algorithmic details).

Acknowledgement

The author wishes to thank Prof. Aitchison of Reading University for drawing his attention to this problem.

References

[1] D. Boley, Computing the controllability/observability decomposition of a linear time-invariant dynamic system, a numerical approach, Ph.D. Thesis, Comp. Sci. Dept., Stanford Univ. (1981).

[2] S. Cumming, Design of observers of reduced dynamics, *Electronic Letters* **5** (1969) 213–214.

[3] G. Golub, S. Nash and C. Van Loan, A Hessenberg–Schur method for the problem $AX + XB = C$, *IEEE Trans. Automat. Control* **24** (1979) 909–913.

[4] B. Gopinath, On the control of linear multiple input–output systems, *Bell System Techn. J.* **50** (1971) 1063–1081.

[5] J. Kautsky, N. Nichols and P. Van Dooren, Robust pole assignment in linear state feedback, *Internat. J. Control*, submitted.

[6] H. Kimura, Geometric structure of observers for linear feedback control laws, *IEEE Trans. Automat. Control* **22** (1977) 846–855.

[7] M. Konstantinov, P. Petkov and N. Christov, Control of linear systems via the serial canonical form, *Proceedings 2nd IFAC/IFIP Symp. Software for Computer Control* (1979) A.II.1–A.II.4.

[8] D. Luenberger, Observing the state of a linear system. *IEEE Trans. Mil. Electr.* **8** (1964) 74–80.

[9] D. Luenberger, Observers for multivariable systems, *IEEE Trans. Automat. Control* **11** (1966) 190–197.

[10] G. Minimis and C. Paige, An algorithm for pole assignment of time invariant multi-input linear systems, *Proceedings 21st IEEE Conf. Decision and Control* (1982) 62–67.

[11] R. Patel, Computation of minimal-order state-space realizations and observability indices using orthogonal transformations, *Internat. J. Control* **33** (1981) 227–246.

[12] G.W. Stewart, *Introduction to Matrix Computations* (Academic Press, New York, 1973).

[13] P. Van Dooren, On the computation of the Kronecker canonical form of a singular pencil, *Linear Algebra Appl.* **27** (1979) 103–141.

[14] P. Van Dooren, The generalized eigenstructure problem in linear system theory, *IEEE Trans. Automat. Control* **26** (1981) 111–129.

[15] P. Van Dooren, A. Emami-Naeini and L. Silverman, Stable extraction of the Kronecker structure of pencils, *Proceedings 17th IEEE Conf. Decision and Control* (1979) 521–524.

[16] P. Van Dooren and M. Verhaegen, The use of condensed forms in linear system theory, *Proceedings IEE Workshop on Reliable Numerical Methods in Control Systems Design*, London (1983) 4/1–4/6.

[17] A. Varga, Numerically stable algorithm for standard controllability form determination, *Electronic Letters* **17** (1984) 74–75.

[18] J.H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).