A reduced order observer for descriptor systems

M.H. VERHAEGEN
NASA Ames Research, M.S. 210-9, Moffet Field, CA 94035, USA

P. Van DOOREN
Philips Research Laboratory, Av. Van Becelaere 2, Box 8, B-1170 Brussels, Belgium

Received 24 February 1986

In this paper a reduced order observer is analyzed for so-called generalized state space (or descriptor) systems. Based on the staircase form for generalized state space models, a recursive algorithm is presented to construct a reduced order observer for a given observable descriptor system. In the algorithm only the eigenfrequencies of the observer have to be specified.

Keywords: Generalized state space model, Reduced order observer, Staircase form, Schur form, Sylvester equation.

1. Introduction

The linear state space model,

$$\lambda x = Ax + Bu, \quad y = Cx$$

(1)

where $x \in R^n$, $u \in R^m$ and $y \in R^p$ and $\lambda$ denotes the differential operator in the continuous-time case and the shift operator in the discrete time case, has proven to be a powerful starting point for the analysis, synthesis and simulation of linear multivariable systems. The state space model (1) results in many practical applications from the linearization of a set of non-linear differential or difference equations around some working point (or reference condition). However, as pointed out by Rosenbrock [5], such a linearization more often gives rise to the following dynamic system representation:

$$\lambda E x = Ax + Bu, \quad y = Cx$$

(2)

by merely writing down the physical relationships. This is e.g. the case in the analysis of systems with switched capacitors [12]. This generalization of the state space model is often called a descriptor system or generalized state space system [11,13].

In order to derive from (2) the state space model (1) is is necessary to invert $E$, which of course can only be done if $E$ is non-singular. Since it may also occur that $E$ is singular, one has to adapt theories and/or algorithms to cope with this more general case as well. In this paper we try to extend results on reduced order observers obtained earlier for standard state space models [2,9] to the case of generalized state space models where $E$ is singular. Nevertheless, the algorithm developed here is also recommended for generalized state space models with non-singular (but poorly conditioned) $E$, this for numerical reasons that will be explained.

2. The reduced order observer problem for descriptor systems

A prerequisite for the design of a reduced order observer is the observability of the system at hand. While for standard state space models this is a well established concept, for descriptor systems there still exists some confusion about this concept when $E$ is singular, i.e. in the presence of so-called infinite frequencies [13]. Here, we will use the following definition derived from [8,16].
Definition 1. A descriptor system of the form (2) is called observable iff the following two conditions hold:

(i) the pencil
\[
\begin{pmatrix}
\lambda E - A \\
C
\end{pmatrix}
\]
has full column rank for all finite values of \( \lambda \),

(ii) the matrix
\[
\begin{pmatrix}
E \\
C
\end{pmatrix}
\]
has full column rank.

Remark 1. Condition (i) of Lemma 1 boils down to observability of the \textit{finite} frequencies of the system (2). It can be retrieved in all current definitions of observability of descriptor systems.

Condition (ii) on the other hand, is where differences occur from author to author, and is related to the observability of the \textit{infinite} frequency.

The above two conditions can also be shown to be equivalent to the geometrical condition used in [8] which is that the \textit{supremal deflating subspace} of \( \lambda E - A \) in the \textit{kernel} of \( C \) has dimension zero [8].

When the descriptor system (2) is observable, as defined above, we now show how to construct a reduced order observer of the form
\[
\begin{align*}
\lambda S z &= F z + D y + P u \\
(5)
\end{align*}
\]
using the input \( u \) and output \( y \) of the original descriptor system (2). The state reconstruction \( \hat{x} \) is then derived from
\[
\begin{pmatrix}
T \\
C
\end{pmatrix} \cdot \hat{x} = \begin{pmatrix} z \\ y \end{pmatrix}
\]
where \( T \) is chosen such that \((T)\) has full column rank.

Remark 2. Without loss of generality we assume in this paper that the matrix \( C \) has linearly independent rows and hence rank \( p \). If this would not be the case, an output transformation could display its linearly independent rows and one would then continue to work with those only.

Remark 3. The word 'state' in the context of descriptor systems is sometimes reserved for that part of vector \( x \) in the \textit{range} of \( E \) [13]. This subtlety is not relevant here though, and we therefore reserve the word 'state' for \( x \) in (2).

It will be shown that an observer (5) of state dimension \( r = (n - p) \) always exists provided the system (2) is observable. The proof is constructive and implicitly provides an algorithm for solving the problem. The development in this section is very similar to that of [2] and [9] used for state space systems.

In order for the state reconstruction \( \hat{x} \) to converge exponentially to the state \( x \) of (2),
\[
\lim_{t \to \infty} (T x - z) = 0,
\]
one obviously needs to choose the unknown matrices \( S, F, D, P \) and \( T \) appropriately. A solution for this is now proposed in terms of an intermediate matrix \( X \).

Theorem 1. Let \( X \) be an \((n - p) \times n\) matrix, such that
\[
\begin{pmatrix}
XE \\
C
\end{pmatrix}
\]
is invertible
\]

(\ref{eq:invertible_matrix})
and satisfying the following ‘generalized’ Sylvester equation:

$$SXA - FXE = DC.$$  \hspace{1cm} (9)

Then putting

$$P = SXB, \quad T = XE,$$  \hspace{1cm} (10)

we have that

$$\lambda S(Tx - z) = F(Tx - z).$$  \hspace{1cm} (11)

Moreover, when the generalized eigenvalues of the observer (5) are asymptotically stable (in the continuous or discrete sense), then exponential convergence of $\hat{x}$ to $x$ occurs.

**Proof.** Multiplying (2a) by $SX$ from the left and substituting (2b), (9), (10) in it, one obtains

$$\lambda S(Tx) = F(Tx) + Dy + Pu.$$  \hspace{1cm} (12)

Comparing this with (5) immediately gives (11). The stability requirement of the observer now implies that its generalized eigenvalues are finite (they lie inside the unit disk for the discrete time case and in the open left half plane for the continuous time case). Hence, $S$ must be invertible and the eigenvalues of $S^{-1}F$ asymptotically stable. One then also finds that the solution $Tx - z$ of (11) converges exponentially to zero. Indeed, we have for the cases of a system of differential equations and difference equations, respectively, that

$$\dot{x}(t) - x(t) = G \cdot e^{S^{-1}Ft} \cdot [z(0) - Tx(0)], \quad \bar{x}(k) - x(k) = G \cdot (S^{-1}F)^k \cdot [z(0) - Tx(0)],$$

where $G$ is defined by

$$(G|H) = \left(\frac{T}{C}\right)^{-1}.$$  \hspace{1cm} (13)

This is easily derived from (6), (11), (14). □

In [2] the first complete solution to the reduced observer problem for a state space model was given, based on the Luenberger canonical form. The idea used there of transforming the problem to a coordinate system where the solution becomes more ‘apparent’ is certainly appealing but also possibly dangerous from a numerical point of view. In the next section we introduce the so-called ‘condensed forms’, which in fact constitute a class of numerically reliable coordinate systems to work in. The freedom of choice is then of course much more restricted but still allows in general to formulate an elegant solution for the problem at hand.

### 3. Condensed generalized state space models

Condensed forms of the system triplet \{A, B, C\} of an ordinary state space model (1) have already been exploited in numerous algorithms because of their appealing property of combining algorithmic efficiency with numerical reliability. An overview of existing condensed forms and a number of new applications are given in [14]. The wide range of applications of condensed forms suggests that these techniques can be considered as a new tool in designing algorithms for system theoretical problems. These forms typically contain as many zeros as possible in the three system matrices. The zero entries are obtained by performing only unitary transformations on the state, input and output spaces:

$$x_1 = Ux, \quad u_1 = Vu, \quad y_1 = Wy.$$  \hspace{1cm} (15)
The corresponding triplet \((A_i, B_i, C_i)\) of the new system has then the following form:

\[
A_i = UAU^*, \quad B_i = UBV^*, \quad C_i = WCU^*
\]  

(16)

where the superscript * denotes the (conjugate) transposed.

For a descriptor system \((2)\), determined by a quadruple of matrices \((A, B, C, E)\), one can use an independent transformation \(Z\) of the equation \((2a)\) in order to obtain an equivalent system in 'condensed' form [10]. The corresponding quadruple \((A_i, B_i, C_i, E_i)\) has then the form

\[
A_i = ZAU^*, \quad B_i = ZBV^*, \quad C_i = WCU^*, \quad E_i = ZEU^*,
\]

(17)

where again all transformations \(U, V, W\) and \(Z\) are unitary. Using these transformations, the following condensed forms can be obtained (among others [10]):

(a) The **generalized Schur form**, where \(A_i\) and \(E_i\) are upper triangular [10,3]. In this form the eigenfrequencies of the system are the ratios \(\alpha_{i,i}/e_{i,i}\) of the corresponding diagonal elements of the triangular matrices \(A_i\) and \(E_i\) (i.e. the generalized eigenvalues of the pencil \(\lambda E_i - A_i\), [3]).

(b) The **generalized (observer) staircase form**, where \(E_i\) is upper block triangular and where \((C_i)\) is upper block-trapezoidal [7,8]. This is illustrated below:

\[
E_i = \begin{bmatrix}
E_{1,1} & E_{1,2} & \ldots & E_{1,k} \\
E_{2,1} & E_{2,2} & \ldots & E_{2,k} \\
\vdots & \vdots & \ddots & \vdots \\
0 & \ldots & \ldots & E_{k,k}
\end{bmatrix}, \quad A_i = \begin{bmatrix}
A_{1,1} & A_{1,2} & \ldots & A_{1,k} \\
A_{2,1} & A_{2,2} & \ldots & A_{2,k} \\
\vdots & \vdots & \ddots & \vdots \\
0 & \ldots & \ldots & A_{k,k}
\end{bmatrix}, \quad C_i = \begin{bmatrix}
0 & \ldots & 0 & C_k
\end{bmatrix}^T
\]

(18)

Here the off-diagonal blocks \(A_{i+1,j}, i = 2, \ldots, k - 1\), and \(C_k\) have full column rank \(r_i\), and the diagonal blocks \(E_{i,i}, i = 1, \ldots, k - 1\), have full row rank \(r_i\), by construction [7,8]. The first off-diagonal block \(A_{2,1}\) is either zero or has full column rank \(r_i\), depending on the termination of the 'staircase' algorithm [7], constructing this form [8].

Under the conditions of observability given in Definition 1, we now have that:

(i) because of condition (i) the block \(A_{2,1}\) has full column rank as well, precluding the existence of unobservable finite frequencies,

(ii) because of condition (ii) the \(E_{i,i}\) are square invertible (hence \(r_i = t_i\)), precluding the existence of unobservable infinite frequencies.

In [4] a further variant is suggested whereby the \(r_i \times r_i\) matrices \(E_{i,i}\) are chosen upper triangular and the \(r_i \times r_{i+1}\) matrices \(A_{i+1,i}\) are chosen upper triangular in the top corner, i.e.:

\[
A_{i+1,i} = \begin{bmatrix}
x & x & \ldots & x & x \\
0 & x & \ldots & x & x \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & x & x \\
0 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

(19)

where the x's are non-zero.
The matrix $E_r$ is thus upper triangular but singular, while its possible singularity always occurs in its last $p$ rows. This is illustrated below for $n = 10$ and $r_1 = 2$, $r_2 = 2$, $r_3 = 3$, $r_4 = p = 3$:

$$A_r = \begin{bmatrix}
xx x x x x x x x x \\
xx x x x x x x x x \\
xx x x x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x
\end{bmatrix}$$

$$E_r = \begin{bmatrix}
xx x x x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x \\
xxxx x x x x x x
\end{bmatrix}$$

$$C_r = \begin{bmatrix}
0 0 0 0 0 0 0 x x x \\
0 0 0 0 0 0 0 0 0 x x \\
0 0 0 0 0 0 0 0 0 0 0 0 0 x
\end{bmatrix}$$

where again the $x$'s are known to be non-zero. This particular form of the system $(A_r, B_r, C_r, E_r)$ is fully exploited later.

These forms are identical to the corresponding condensed forms of the standard state space model (1), as e.g. defined in [10], except for $E$, which is upper triangular in each case. Algorithms have been proposed in the literature [4] to obtain these 'generalized' condensed forms. Unfortunately, these algorithms have a complexity $O(n^4)$, while comparable algorithms for the ordinary state space model (1) have complexity $O(n^3)$. Fast $O(n^3)$ implementations to obtain the above staircase forms have now been derived by using techniques used e.g. in [3] and [14], and variants are still under investigation [1]. These fast algorithms will not be discussed here although they make the ideas presented in this paper more appealing.

Remark 4. If the system model $(A, B, C, E)$ is real, it will have a complex generalized Schur form when the generalized eigenvalues of $\lambda E - A$ are complex. In this case one may prefer to construct the real generalized Schur form [3] where $A_r$ has $2 \times 2$ bumps on its diagonal, each block corresponding to a pair of complex conjugate eigenvalues. The $E_r$ matrix is still triangular and real as well. The advantage of this form is that one maintains all transformations and subsequent computations in real arithmetic, which e.g. may result in a significant speed-up.

Remark 5. For any of the above condensed forms there exist dual forms where 'upper' is replaced by 'lower'. One can thus define a lower generalized Schur form and a lower generalized staircase form, which will then be a controller staircase form [10]. The lower variant of the Schur form will be used in the sequel together with the (upper) observer staircase form.

Following the lines of [9], we choose here a coordinate system for both $(A, B, C, E)$ defining the system (2), and for $(S, F, D, P)$ defining the observer (5), such that (2) is in observer staircase form as described in (18)-(20) and (5) is in lower Schur form. For the system (2) this thus implies a preliminary
transformation of the type (17) where $V$ can be chosen the identity [8]. A similar transformation could be performed for the observer (5). However, since these system matrices still have to be determined, their structure can be chosen and therefore no explicit transformation is required.

The equations (8)-(10) are now reduced to the equivalent equations

$$SX_iA_i - FX_iE_i = D_iC_i \quad \text{with} \quad \begin{pmatrix} X_iE_i \\ C_i \end{pmatrix} \text{invertible,}$$

$$P_i = SX_iB_i, \quad T_i = X_iE_i \quad \text{(22)}$$

where the relations between the transformed solutions (indexed by $i$) and those of Theorem 1 are

$$T_i = TU^*, \quad X_i = XZ^*, \quad P_i = P, \quad D_i = DW^*. \quad \text{(23)}$$

It is important to note here that by this specific choice of unitary transformations, we obtain an equivalent numerical problem as well. By this we mean that the sensitivities of the solution for $T_i$, $X_i$, $P_i$ and $D_i$ are the same as for $T_1$, $X_1$, $P_1$ and $D_1$, the reason for this being the unitarity of $U$, $Z$ and $W$ (see also [8,10]).

4. A recursive algorithm

In this section we derive a recursive solution for the equation (21). We only assume $E_1$, $A_1$, $C_1$ and the generalized eigenvalues of the reducible order observer (2) to be given, while $S$, $F$, $D_i$ and $X_i$ are to be constructed. We propose a solution for $X_i$ in the form

$$X_i = \begin{pmatrix} 1 & x'_1 \\ 0 & 1 \\ 0 & 0 & 1 & x'_2 \\ \vdots & \vdots \\ 0 & 0 & 0 & 1 & x'_r \end{pmatrix}, \quad (24)$$

where only the rows $x'_i$, $i = 1, \ldots, r$, of length $n - i$ have to be determined. The reason for choosing this form is that because of this and (20), the matrix $(X_i^{X_i'})$ is then upper triangular and invertible. One then only has to solve the Sylvester equation (21a) to derive a solution (21), (22) for the problem. This is done as follows. Let us write the matrices $S$, $F$ and $D$ similarly to $X_i$:

$$S = \begin{pmatrix} \mu_1 & 0 & 0 & 0 & 0 \\ s'_1 & \mu_2 & 0 & 0 & 0 \\ s'_2 & \mu_3 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s'_{r-1} & \mu_r \\ s'_{r} & \mu_1 & 0 & 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} \nu_1 & 0 & 0 & 0 & 0 \\ f'_1 & \nu_2 & 0 & 0 & 0 \\ f'_2 & \nu_3 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f'_{r-1} & \nu_r \\ f'_{r} & \nu_1 & 0 & 0 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} d'_1 \\ \vdots \\ \vdots \\ d'_{r-1} \\ d'_{r} \end{pmatrix}. \quad \text{(25)}$$

Here the diagonal elements of $S$ and $F$, denoted by $(\mu_i)$ and $(\nu_i)$ respectively, are given since their ratio's $\mu_i/\nu_i$ are the generalized eigenvalues of the observer (5).

**Remark 6.** Any solution $S$, $F$, $D_i$ to (21a) can be arbitrarily scaled to yield another solution. A normalization is automatically introduced by fixing the diagonal elements $(\mu_i)$ and $(\nu_i)$. Since only their ratio's are imposed we could choose them such that e.g. $\mu_1^2 + \nu_1^2 = 1$. This automatically fixes the freedom of choice in these diagonal elements.
Using the particular structure (24) and (25), the first row of (21) can be rewritten as

\[(x'_1|d'_1) \cdot \begin{pmatrix} A_{\mu r} & -C_r \end{pmatrix} = -a_{\mu r}, \]  

(26)

where \(A_{\mu r}\) and \(a_{\mu r}\) denote respectively the bottom \(n-1\) rows and the first row of the matrix \(\mu_1 A_i - \nu_1 E_i\). Similarly we rewrite the \(i\)-th row of (21) as

\[
\begin{pmatrix} s'_i & f'_i & X'_i & d'_i \\ i-1 & i-1 & n-i & p \end{pmatrix} \begin{bmatrix} X_i A_i \\ -X_i E_i \\ A_{\mu r} \\ -C_i \end{bmatrix} = -a_{\mu r},
\]

(27)

where \(A_{\mu r}\) and \(a_{\mu r}\) denote respectively the bottom \(n-i\) rows and the \(i\)-th row of the matrix \(\mu_i A_i - \nu_i E_i\) and \(X_i\) denotes the top \(i-1\) rows of \(X_i\). Equations (26) and (27) can be denoted as a system of equations of the type

\[X'_i M_i (X_1, \ldots, X_{i-1}) = a_{\mu r}, \]

(28)

in the unknown \(X'_i = (s'_i | f'_i | x'_i | d'_i)\), whereby \(M_i\) depends on the previous vectors \(X'_i\). One can thus solve recursively for the rows of \(S, F, X,\) and \(D\), using (26) and (27) for \(i = 1, \ldots, n-p\) provided all of the \(M_i\) matrices have a left inverse. This now easily follows from the particular structure (20) of the condensed form we chose to start with. Indeed, a typical configuration of an \(M_i\) matrix for a system with stairs \(r_1 = 2, r_2 = 2, r_3 = 3, p = r_4 = 3\) (the same situation as in (20)) would be, say at step \(i = 3\),

\[
M_i = \begin{bmatrix} x & x & x & x & x & x & x & x \\ x & x & x & x & x & x & x & x \\ x & x & x & x & x & x & x & x \\ 0 & x & x & x & x & x & x & x \\ 0 & 0 & x & x & x & x & x & x \\ 0 & 0 & 0 & x & x & x & x & x \\ 0 & 0 & 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{i-1}
\]

\[
\begin{bmatrix} x & x & x & x & x & x & x & x \\ x & x & x & x & x & x & x & x \\ x & x & x & x & x & x & x & x \\ 0 & x & x & x & x & x & x & x \\ 0 & 0 & x & x & x & x & x & x \\ 0 & 0 & 0 & x & x & x & x & x \\ 0 & 0 & 0 & 0 & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{n-i}
\]

(29)

Because of the leading non-zero x's in each column, these \(M_i\) matrices will always have full column rank, independently of the choice of generalized eigenvalues for the observer. These systems are in fact underdetermined and can be solved for \(X_i\) under the constraint that the norm of \(X_i\) should be minimal. This choice is a logical one since it minimizes the norm of the matrix \(D\), and the off diagonal part of \(F, S\) and \(X_i\). These minimal norms have a positive effect on the condition number for the generalized eigenvalue problem of \(\lambda S - F\) and of the inversion of the compound matrix (21b), two numerical properties that are welcome in this problem, see also [9]. In order to construct the 'minimum norm' solution of (26) and (27), one performs a \(QR\) decomposition of the matrices \(M_i\), [6,15]:

\[M_i = Q_i R_i. \]

(30)
Using Householder transformations, this requires approximately \((p + i)n^2\) 'flops' (1 flop = 1 multiplication + 1 addition) per \(M_i\) matrix. Indeed, because of the use we made of the observer staircase form, these matrices \(M_i\) (as illustrated in (24)) have at most \(p + i - 3\) non-zero subdiagonals and can then be decomposed with relatively few operations. The computation of the solution of the triangular system \(R_i\), and the back transformation with \(Q_i\), are negligible with respect to the decomposition. The total cost is (roughly) given by

\[
\frac{1}{2}(n - p)(n + p)n^2 \text{ flops for solving for } S, \ F, \ D, \ \text{and } X_i.
\] (31)

For the construction of \(P\) and \((T_i)^{-1}\) we use

\[
P = SX_iB_i \quad \text{and} \quad \begin{pmatrix} T_i \\ C_i \end{pmatrix}^{-1} = U' \begin{pmatrix} T_i \\ C_i \end{pmatrix}^{-1} \begin{pmatrix} I_r \\ 0 \\ 0 \\ W \end{pmatrix},
\] (32)

which requires (using the special structure of \(S, X_i, T_i, \text{ and } C_i\)) approximately

\[
(n - p)nm + \frac{1}{2}(n - p)^2n \quad \text{and} \quad \frac{1}{2}n^3
\] (33)

flops, respectively. The major part of the computations thus goes in the decomposition of the \(M_i\) matrices which may require \(O(n^4)\) operations. Yet, if one is happy to drop the minimum norm condition, one can easily obtain a solution for (26), (27) by e.g. merely selecting from each \(M_i\) those rows that have a non-zero leading element \(x\) and solving this subsystem. This would then become an \(O(n^3)\) algorithm for the complete solution of the problem (see also the remarks in Section 3 on the construction of the condensed forms).

**Remark 7.** When the observer has complex eigenvalues and we restrict to real computations, a complete analogous procedure can be developed as the one given above. This is very similar to the technique described for ordinary state space models in [9] and therefore we refer to this paper.

5. Conclusion

In this paper we showed that the use of generalized condensed forms can lead to efficient algorithms for descriptor systems in much the same way as algorithms based on condensed forms do for ordinary state space models (it is interesting to notice that for \(E = I\) and \(S = I\) the equivalent method developed in [9] is indeed retrieved). Although the method is developed specifically for singular \(E\), we also recommend its use for non-singular \(E\). The possibility to reduce the descriptor system (2) to a standard state space model (1) should indeed be avoided if possible due to the possible loss of accuracy that is often incurred by this step (when \(E\) is badly conditioned with respect to inversion). The present technique in fact implicitly solves the same problem, but without passing via a standard state space model and the possible loss of accuracy connected to this.

**References**


