Systems & Control Letters 2 (1982) 202–208
North-Holland Publishing Company

December 1982

Approximate realization based upon an alternative to the Hankel matrix: the Page matrix

A.A.H. DAMEN, P.M.J. Van den HOF

Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

A.K. HAJDASINSKI

Samenwerkingsorgaan KHT–THE, Catholic University Tilburg, The Netherlands

Received 27 July 1982
Revised 16 October 1982

The Ho–Kalman algorithm creates a minimum realization of a system, when given a series of deterministic Markov parameters. However, when such a 'truncated' series of Markov parameters has been disturbed with noise, an approximating Hankel matrix has to be constructed for applying the realization algorithm. This approximating Hankel matrix has either the improper rank, or it lacks the Hankel structure.

Furthermore the Markov parameters are not processed with a constant weighting factor, which implies that the noise filtering is inadequate. In this paper we propose to use an alternative matrix: the Page matrix. It is shown that this method is better suited for handling the noisy Markov parameters. This holds with respect to three aspects: order testing, noise filtering and realization.

Even in the deterministic case, the Page matrix offers the advantage of a considerable reduction in computation.

Keywords: Multivariable systems, Stochastic systems, Identification, Parameter estimation, System order reduction, Noise filtering, Realization, Hankel matrix.

Introduction

The minimum realization for a sufficiently long truncated series of deterministic Markov parameters offers no problems. This has been shown by Ho and Kalman [1]. In Section 1 we will summarize this algorithm. However, a fully satisfactory solution for the noisy case has not yet been proposed. Of course, the Ho–Kalman algorithm is being applied in a modified way for the noisy case, but the results are rather questionable. This is quite understandable because, theoretically, the processing of the noise is fundamentally wrong. This inadequacy will be elucidated in Section 2. In our opinion, the use of the Page matrix, which will be introduced in Section 3, may overcome the majority of the problems caused by the use of the Hankel matrix. This will be emphasized in Section 4, where we will compare the Hankel matrix algorithm with the one based on the Page matrix.

1. The Ho–Kalman algorithm for deterministic data

As a short recapitulation and in order to define a notation, we will briefly sketch the Ho–Kalman algorithm [1]. It constructs a minimum realization (A, B, C) of a linear, time-invariant, state space model, given a noise-free Hankel matrix with the correct size for the system.

Let a truncated series of Markov parameters be denoted by

$M_1, M_2, M_3, \dots, M_L$

$$\text{where}\quad \begin{aligned} &M_i = CA^{i-1}B \quad (L \text{ even}) \\ &\dim(M_i) = q \times p, \\ &q = \text{number of outputs}, \\ &p = \text{number of inputs}. \end{aligned} \tag{1}$$

Then a Hankel matrix and its decomposition can be written as

$$\begin{aligned} H &= \begin{bmatrix} M_1 & M_2 & M_3 & \dots & M_{L/2} \\ M_2 & M_3 & M_4 & \dots & M_{L/2+1} \\ \vdots & & & & \vdots \\ M_{L/2} & & & \dots & M_{L-1} \end{bmatrix} \\ &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L/2-1} \end{bmatrix} \cdot \begin{bmatrix} B & AB & A^2B & \dots & A^{L/2-1}B \end{bmatrix} \\ &= \Gamma \cdot \Delta. \end{aligned} \tag{2}$$

Suppose we are dealing with a completely observable and completely controllable system. If $L/2 \geqslant n$, the minimum dimension of the system (sufficient condition [1]), the rank of H will equal n. Any such decomposition into Γ^* and Δ^* of minimum dimension n and full rank will show the same structure as in formula (2) with a corresponding minimum realization (A^*, B^*, C^*), because

$$H = \Gamma^* \Delta^* = \Gamma\Delta, \tag{3}$$

implying

$$\Delta^* = (\Gamma^*)^+ \Gamma\Delta = T\Delta \tag{4}$$

where $+$ stands for pseudo inverse and T is nonsingular as Γ and Γ^* are of full rank n. Analogously we may put

$$\Gamma^* = \Gamma\Delta(\Delta^*)^+ = \Gamma S. \tag{5}$$

Finally, substitution of (4) and (5) into (3) leads to

$$H = \Gamma\Delta = \Gamma S T \Delta \tag{6}$$

so that

$$\Gamma^+ \Gamma S T \Delta \Delta^+ = I \;\Rightarrow\; T = I \;\Rightarrow\; S = T^{-1}. \tag{7}$$

So the equivalence transformation can be defined as

$$\left.\begin{array}{ll} \Delta^* = T\Delta, & \Gamma^* = \Gamma T^{-1}, \\ B^* = TB, & C^* = CT^{-1}, \\ A^* = TAT^{-1}. & \end{array}\right\} \tag{8}$$

The complete set of all possible (Γ, Δ) together then produces the complete equivalence class of the system under study.

The triplet (A, B, C) can be obtained from (Γ, Δ) as follows: the matrices B and C can be recognized as the first blocks in Δ and Γ respectively. In order to obtain matrix A, we need a shifted matrix, which we indicate by an arrow. A vertically pointing arrow indicates a shift of one block row, whereas a horizontally pointing arrow denotes a shift of one block column. From this it is clear that

$$H\uparrow = \overleftarrow{H} = \Gamma A \Delta \tag{9}$$

and as Γ and Δ have a maximum rank n, we may write

$$A = \Gamma^+ H \uparrow \Delta^+. \tag{10}$$

Note that we needed the extra Markov parameter M_L to construct the shifted Hankel matrix.

An alternative procedure is provided by either the 'extended observability' matrix Γ or the 'extended controllability' matrix Δ, because

$$\Gamma\uparrow = \Gamma A \quad \text{or} \quad A\Delta = \dot{\Delta}. \tag{11}$$

Consequently this yields

$$A = \Gamma^+ \Gamma^{\uparrow} \quad \text{or} \quad A = \dot{\Delta}\Delta^+. \tag{12}$$

In order to construct the shifted matrices $\Gamma\uparrow$ or $\dot{\Delta}$ having the same dimensions as Γ or Δ, we lack information on what to insert in the latter blocks of Γ or Δ during this operation. Therefore we have to apply a Γ and a Δ with reduced dimensions in (11) and (12). The reduction is accomplished by omitting the last block.

Finally, a numerically stable decomposition of H is offered by the singular value decomposition, given by

$$H = UDV^{\mathrm{T}} \quad \dim(H) = g \times l, \tag{13}$$

where

$$D = \mathrm{diag}(\delta_1, \delta_2, \delta_3, \ldots, \delta_s),$$
$$\delta_1 \geqslant \delta_2 \geqslant \delta_3 \geqslant \cdots \geqslant \delta_n > 0,$$
$$\delta_{n+1} = \delta_{n+2} = \cdots = \delta_s = 0,$$
$$s = \min(g, l),$$
$$U^{\mathrm{T}}U = I_s \quad \text{and} \quad V^{\mathrm{T}}V = I_s.$$

Because $\mathrm{rank}(H) = n$ we may rewrite this as

$$H = H_n = U_n D_n V_n^{\mathrm{T}} \tag{14}$$

where for H_n only the first n nonzero singular values are used in D_n, and the corresponding singular vectors (i.e. columns) in U_n and V_n.

If we distribute the singular values in a balanced way among Γ and Δ we get

$$\Gamma = U_n D_n^{1/2} \quad \text{and} \quad \Delta = D_n^{1/2} V_n. \tag{15}$$

These matrices have the proper dimensions and full rank n, which is sufficient to guarantee the proper structure according to formula (2) for the deterministic case, as shown in formulae (3) to (8).

[1] As is generally known, the necessary lower bounds are given by the observability and controllability indices, while non-block-symmetric Hankel matrices may be used.

2. Application of the Ho–Kalman algorithm for noisy data

In 1974 Zeiger and McEwen [2] suggested using the Ho–Kalman algorithm with the singular value decomposition in the case of noisy data. All singular values up to and including δ_s will then be nonzero. (The distinction with the deterministic case is denoted by the circumflex $\hat{}$.)

It is easy to verify that, in cases where the Markov parameters are contaminated with *Swaying* noise, we may write (see appendix)

$$E\{\hat{\delta}_i^2\} = \delta_i^2 + \sigma^2 \max(l, g), \tag{16}$$

where E stands for expectation.

Swaying noise is defined as follows:

$$\hat{M}_i(a, b) = M_i(a, b) + \xi_{iab}, \tag{17}$$

where $M_i(a, b)$ is element a, b in matrix M_i and ξ_{iab} is the corresponding additive noise. This noise is assumed to be stationary (S), white (W) (zero mean), additive (A), signal-independent (Y), interindependent (among channels) (I), with nonchanging global variance σ^2 (NG):

$$E\{\xi_{iab}\} = 0 \quad \forall i, a, b \in \mathbb{N}, \tag{18}$$

$$E\{\xi_{iab}\xi_{jcd}\} = \begin{cases} 0 & i, a, b \neq j, c, d, \\ \sigma^2 & i, a, b = j, c, d. \end{cases} \tag{19}$$

This increase in singular values is reflected in Fig. 1.

Based upon knowledge of the noise level in the singular value, one may decide upon the dimension n of the system and approximate the Hankel matrix in a least squares sense by

$$H_n = U_n D_n V_n^{\mathrm{T}} \tag{20}$$

(see e.g. [3]).

In this way we perform a noise filtering on the Hankel matrix and implicitly use the singular values for the order testing, before we apply the Ho–Kalman algorithm.

Then, there exist several possibilities for applying the formulae of the previous section in order to find a realization. These possibilities are compared in [4].

For this shifted Hankel matrix we may either use the original H [5], or the approximating H_n [6]. Alternatively, the Γ and Δ matrices may be used [7] to evaluate A.

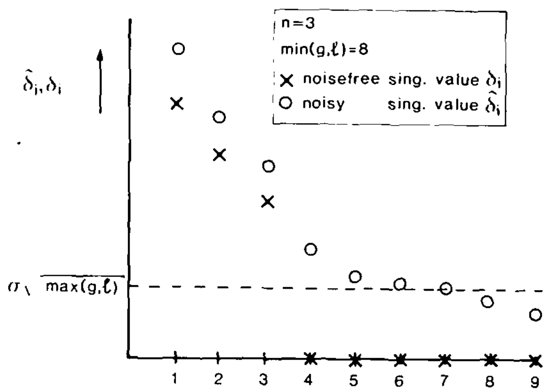

Fig. 1. Singular values of the Hankel matrix.

All these possibilities lead to a suboptimal realization because the obtained Γ and Δ matrices do not have the proper structure of extended observability and extended controllability matrices according to formula (2). This is due to the remaining noise influence which is not filtered out adequately in the first step: the approximation of the noisy Hankel matrix by a proper rank matrix.

Although the approximating Hankel matrix H_n may have the proper rank, it lacks the necessary Hankel structure. Therefore it does not provide directly unique Markov parameters. From another point of view, the Markov parameters, that constitute the Hankel matrix, are weighted by an isosceles triangular function with the top on $L/2$ for $M_{L/2}$ (see Fig. 2). This is due to the fact that, depending on their index, the Markov parameters appear more frequently in the Hankel matrix.

Nevertheless, this (noise adapted) Ho–Kalman algorithm weights *all* available Markov parameters at least with a *nonzero* weighting factor. If we were to use the algorithm suggested by Silverman [8], we would select a full rank submatrix from the Hankel matrix, which is just big enough to calculate a realization (see Fig. 2). Because a strictly limited part (first) of the Markov sequence is used, this is quite inappropriate for noisy cases (no redundancy): the basis, thus obtained for the realization, may be quite ill-conditioned numerically.

Going back to the Ho–Kalman algorithm, we note that statistical considerations are also difficult to make, as the noise on the entries in the Hankel matrix is not independent, but exactly the same noise data appear frequently in several entries.

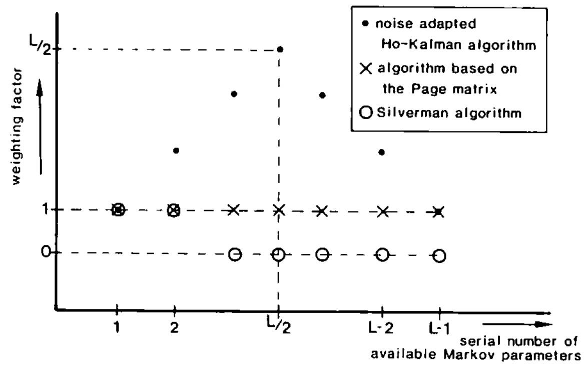

Fig. 2. Weighting of the Markov parameters in the least squares approximation of the Hankel matrix or Page matrix, due to the number of appearances in those matrices. In the case of the noise adapted Ho–Kalman algorithm we get an isosceles triangular weighting. For the Page matrix algorithm the pursued constant weighting is accomplished. In the Silverman algorithm just the first Markov parameters are used until the proper rank is obtained. Not all elements in these Markov parameters are necessarily used and all information in the neglected remaining Markov parameters is lost.

Because of all these drawbacks when using the Hankel matrix, we will now introduce an alternative matrix.

3. Introduction of the Page matrix

In order to overcome the problems caused by the special block structure of the Hankel matrix, a trivial matrix is introduced here, which is constructed from the Markov parameters in the most natural way.

Similarly to filling a page with characters, it follows that this matrix should be called the Page matrix. It is defined as

$$P=\begin{bmatrix} M_1 & M_2 & M_2 & \dots & M_\mu \\ M_{\mu+1} & & & \dots & M_{2\mu} \\ \vdots & & & & \vdots \\ M_{(\eta-1)\mu+1} & & & \dots & M_{\eta\mu} \end{bmatrix},$$

$$\dim(P)=h\times m. \tag{21}$$

As ever, some nations show a different behaviour in this respect, and consequently we may also define a Chinese Page matrix C_p, where the Markov parameters are ordered column-wise.

We will not analyse this Chinese Page matrix meticulously here, as this analysis is completely dual to the normal Page matrix.

In the purely deterministic case, the Page matrix can be decomposed in a manner similar to the Hankel matrix:

$$P=\begin{bmatrix} C \\ CA^\mu \\ \vdots \\ CA^{(\eta-1)\mu} \end{bmatrix}\cdot\begin{bmatrix} B & AB & A^2B & \dots & A^{\mu-1}B \end{bmatrix}$$
$$=\Gamma_\mu\cdot\Delta. \tag{22}$$

Whereas the Hankel matrix is the product of the extended observability matrix Γ and the extended controllability matrix Δ of the system (A, B, C), the Page matrix is the product of the extended observability matrix of the system (A^μ, B, C) and the extended controllability matrix of the system (A, B, C).

Next we will state the crucial theorem in this context:

Theorem. *If the dimensions of the Page matrix are chosen large enough, then $\mu, \eta \geqslant n$ ($=$ the dimension of the system) is sufficient, and if (C, A^μ) is a completely observable couple, it holds that*

$$\operatorname{rank} P = n \tag{23}$$

and any decomposition in Γ_μ and Δ of minimum dimension n will lead to a minimum realization according to formulae (9)–(12).

The formal proof of this theorem is straightforward and given in [10].

Furthermore in [10] all conditions for the observability of the couple (C, A^μ), which is crucial, are stated. As a short outline we give the following summary.

Since (C, A) is an observable pair, a *sufficient* condition for the observability of (C, A^μ) is given by Hautus in Theorem III of [9]:

$$\phi'(\lambda)\neq 0, \qquad \lambda\in\sigma(A), \tag{24a}$$
$$\phi(\lambda_i)\neq\phi(\lambda_j), \quad \lambda_i, \lambda_j\in\sigma(A), \lambda_i\neq\lambda_j, \tag{24b}$$

where $\phi(\lambda)=\lambda^\mu$ and $\sigma(A)$ is the set of eigenvalues of A.

The *necessary* condition, however, appears to be less severe. In the cases where (a) or (b) happen, (C, A^μ) can be transformed into a Jordan canonical form and then the columns in C_J, corresponding to coinciding poles ($\lambda_i^\mu = \lambda_j^\mu$) and those corresponding to zero poles ($\mu\lambda^{\mu-1} = 0$), have to form an independent set.

The most frequent cause of non-observability of (C, A^μ) will be multiple poles in the origin of the z-plane. These occur in the case of delays and finite impulse responses. This is discussed and elaborated upon in [10].

Because exclusively completely observable and controllable systems (A, B, C) are considered here, non-observability of the system (A^μ, B, C) due to violation of condition (b) will be quite exceptional. It happens if distinct poles in A happen to be non-distinct ones in A^μ.

Let the position of a pole of system matrix A in the complex z-plane be characterized by the radius r and the argument ϕ. Then the corresponding pole for system matrix A^μ is characterized by r^μ and $\mu\phi$. Consequently, original poles of equal radius may coincide in system matrix A^μ in the case of $\mu\phi_i = \mu\phi_j + 2k\pi$, $k \in \mathbb{Z}$. Then the non-observability might occur if the corresponding columns of C_J are dependent, where (C_J, A_J^μ) is the Jordan canonical form of (C, A^μ).

It is obvious that this can be avoided by requiring that for all poles $\mu|\phi| < \pi$ holds, which, for example, implies a sufficiently high sampling rate.

From here we will assume (C, A^μ) to be completely observable. In the dual case of the Chinese Page matrix (which could be used in cases of a suspected failure of the Page matrix), the assumption of complete controllability of (A^μ, B) is made.

Under these conditions the rank of the Page matrix simply defines the order, and the realization (A, B, C) can be obtained similarly as for the Hankel matrix, apart from the fact that the proper shifted matrices have to be used. The Page matrix has to be shifted one block to the left, and the Chinese Page matrix one block upwards. For the use of formulae (12) the Page matrix provides the proper Δ and the Chinese Page matrix the proper Γ.

4. A comparison of the Hankel and Page matrix

In the deterministic case, both the Hankel and the Page matrices can be used to obtain a minimum realization.

As indicated, the Page matrix may fail in some (exceptional) cases. The superiority of the Page matrix is significant for the noisy case. As we have to eliminate the noise, a long Markov sequence (of fixed number L) has to be used. In that case the size of the Page matrix is much smaller ($hm = (L-1)pq$) than the size of the hankel matrix ($lg = pqL^2/4$). Consequently, the reduction of the computational effort is considerable. In the Page matrix all Markov parameters appear only once, which means that, when reducing the rank with the aid of singular value decomposition, there is an equally balanced filtering over the Markov parameters (see Fig. 2). Moreover, a Page matrix of reduced rank directly provides a unique sequence of Markov parameters contrary to the approximated Hankel matrix.

So the noise filtering by means of the singular value decomposition of the Page matrix is simply a least squares approximation of the Markov parameters with a fixed (or estimated) dimension of the system. This dimension has then to be understood as the rank of the Page matrix. In the noisy case the Γ_μ en Δ matrices still do not have the proper structure as defined in formula (2), which was also a drawback when using the Hankel matrix approach.

This noise filtering operation provides us with a criterion for the optimum size of the Page matrix.

For a given number L of available Markov parameters, the block dimensions of the Page matrix can be chosen in different combinations as long as $\eta\mu = L - 1$. If we assume that the Markov parameters in the Page matrix are disturbed with Swaying noise, the total noise energy in P (in expectation) can be written as:

$$hm\sigma^2. \tag{25}$$

Because of the character of the noise, the expected noise energy will, if applying the singular value decomposition to P, be equally distributed over all squared singular values (see Appendix). When the rank of the Page matrix is reduced to n, by setting $\min(h, m) - n$ singular values equal to zero, the expected noise reduction equals

$$\frac{n \max(h, m)\sigma^2}{hm\sigma^2} = \frac{n}{\min(h, m)}. \tag{26}$$

For optimizing this noise reduction, $\min(h, m)$ has to be maximized. This implies that we have to choose P as close to square as possible.

Conclusions

A Page matrix has been proposed as an alternative to the Hankel matrix in the realization problem.

The Page matrix is especially superior in the noisy case in three aspects:

- The order testing: the decision concerning the dimension of the system, based on the singular values of the Page matrix, is straightforward, since all noisy data appears only once in the Page matrix. In cases of Swaying noise, the non-relevant singular values are independent, which is not the case for the Hankel matrix (see Appendix).
- The noise filtering by omitting the non-relevant singular values: there is a constant weighting factor for the Markov parameters and the total reduction equals $n/\min(h, m)$, which is optimal for a square Page matrix. The size of the Page matrix may be chosen smaller than the size of the Hankel matrix for a fixed number L of Markov parameters. This implies a reduction in computation.
- The approximate realization: the noise filtering provides us directly with a set of unique Markov parameters in the approximated Page matrix of rank n, contrary to the situation for the Hankel matrix. This proves that we have reduced the information to the proper degree of freedom by using the Page matrix. Then the realization is straightforward, as in the deterministic case. In the approximated Hankel matrix H, however, a number of superfluous degrees of freedom is still implicitly incorporated.

A possible drawback of the Page matrix for the noisy case can be traced back to the conditions $\lambda_i \neq 0$ and $\lambda_i^\mu \neq \lambda_j^\mu$ for the deterministic case. In the presence of noise, dependent on the rate of deterioration, it will be hard to distinguish whether the behaviour of the system is due to small eigenvalues or zero eigenvalues and likewise to multiple or neighbouring eigenvalues. So in this respect we can expect difficulties, which will be checked in simulations.

Preliminary practical tests confirm the above theoretical expectations and we hope to present these results in a subsequent paper.

Finally, we are also optimistic about the use of the Page matrix for the stochastic realization, where estimates of covariances are replacing the Markov parameters. Here the problem is that the uncertainties of the estimated covariances are far from independent and stationary. On the other hand, zero eigenvalues which were the cause of some problems, cannot happen.

Appendix

The statistical behaviour of the singular values of a Hankel matrix, built up by noise corrupted Markov parameters, is a special case of the situation where all elements of a matrix have independent, additive noise, like the Page matrix. In the Hankel matrix the Markov parameters appear repeatedly and thus the noise cannot be approached as being independent. Nevertheless this independency makes it easy to study the behaviour so that we will first study the noise contaminated Page matrix.

Assume a Page matrix P with dimensions $h \times m$, composed of $L-1$ deterministic Markov parameters, all disturbed by *Swaying* noise with variance σ^2 (see Section 2). The deterministic Markov parameters construct a deterministic Page matrix P, so the following can be written:

$$\hat{P} = P + \Xi p$$

where Ξp is the matrix containing all noise samples.

If $h \leqslant m$ we continue with $\hat{P}\hat{P}^{\mathrm{T}}$; in the opposite case a dual version may be derived by means of $\hat{P}^{\mathrm{T}}\hat{P}$.

$$\hat{P}\hat{P}^{\mathrm{T}} = PP^{\mathrm{T}} + \Xi_P \Xi_P^{\mathrm{T}} + \Xi_P P^{\mathrm{T}} + P\Xi_P^{\mathrm{T}}.$$

Because of the character of the noise, $E(\Xi_p)$ will be zero, and

$$E\{\Xi_p \Xi_p^{\mathrm{T}}\} = m\sigma^2 I_h.$$

As a result,

$$E\{\hat{P}\hat{P}^{\mathrm{T}}\} = PP^{\mathrm{T}} + m\sigma^2 I_h.$$

We can state that $P = UDV^{\mathrm{T}}$, and also

$$m\sigma^2 I_h = U(m\sigma^2 I_h)U^{\mathrm{T}}$$

because of the diagonal character of this matrix and the orthonormality of U.

Therefore we can write

$$E\{\hat{P}\hat{P}^{\mathrm{T}}\} = U(D^2 + \sigma^2 m I_h)U^{\mathrm{T}}.$$

Generally, for all possible (h, m) a description can be given for the matrix $\hat{D}$, the diagonal matrix of the singular values of $\hat{P}$,

$$E\{\hat{D}^2\} = D^2 + \sigma^2 \max(h, m) I_{\min(h,m)}.$$

Although for the Hankel matrix the noise elements appear more frequently, the structure is such that it does not violate the steps used above. So the same conclusion can be made concerning the Hankel matrix. Note, however, that for the Hankel matrix the actual non-relevant singular values are highly dependent.

References

[1] B.L. Ho and R.E. Kalman, Effective construction of linear state-variable models from input/output functions. *Regelungstechnik* 14 (1966) 545–548.

[2] H.P. Zeiger and A.J. McEwen, Approximate linear realizations of given dimension via Ho's algorithm. *IEEE Trans. Automat. Control* 19 (1974) 153.

[3] J. Staar, J. Vandewalle and M. Wemans, Realization of truncated impulse response sequences with prescribed uncertainty, in: H. Akashi, Ed, *Control Science and Technology for the Progress of Society*, Proc. 8th Triennial World Congress of IFAC, Kyoto, 24–28 August 1981, Vol. 1 (Pergamon, Oxford, 1982).

[4] A.A.H. Damen and A.K. Hajdasinski, Practical tests with different approximate realizations based on the singular value decomposition of the Hankel matrix, in: G.A. Bekey, Ed., *Identification and System Parameter Estimation*, Preprints 6th IFAC Symp., Arlington, Virginia, 7–11 June 1982 (Pergamon, Oxford, 1982) pp. 809–814.

[5] G.A. van Zee and O.H. Bosgra, The use of realization theory in the robust identification of multivarible systems, in: R. Isermann, Ed., *Identification and System Parameter Estimation*, Proc. 5th IFAC Symp., Darmstadt, 24–28 Sept. 1979 (Pergamon, Oxford, 1980) pp. 477–484.

[6] A.K. Hajdasinski and A.A.H. Damen, Realization of the Markov parameter sequences using the singular value decomposition of the Hankel matrix, Report 79-E-95, Department of Electrical Engineering, Eindhoven University of Technology (1979).

[7] S. Kung, A new identification and model reduction algorithm via singular value decompositions, in: C.-C. Hsieh, Ed., *Proc. 12th Asilomar Conf. on Circuits, Systems and Computers*, Pacific Grove, Calif., 6–8 Nov. 1978 (IEEE, New York, 1979) pp. 705–714.

[8] L.M. Silverman, Realization of linear dynamical systems, *IEEE Trans. Automat. Control* 16 (1971) 554–567.

[9] M.L.J. Hautus, Controllability and observability conditions of linear autonomous systems. *Proc. Kon. Akad. Wet. Ser. A* 72 (= *Indag. Math.* 31) (1969) 444–448.

[10] A.A.H. Damen, P.M.J. Van den Hof and A.K. Hajdasinski, The Page matrix: an excellent tool for noise filtering of Markov parameters, order testing and realisation, Report 82-E-127, Dept. Elec. Eng., Eindhoven Univ. Technol. (1982).