Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline Conference Paper Archive

IFAC PapersOnLine 56-2 (2023) 5837–5842

Order Selection for Stochastic Bilinear Systems ⋆

József Bokor * **György Terdik** **

* *Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary (e-mail: bokor@sztaki.hu)*

** *Faculty of Informatics, University of Debrecen Debrecen, Hungary (e-mail: terdik.gyorgy@inf.unideb.hu)*

Abstract: Identification of bilinear systems subject to white inputs is studied. Assuming bilinearity we use the crosscovariances between the output and the higher order Hermite polynomials of the input for estimating the coefficients of the Hermite series expansion of the output. The parameters of the bilinear model are obtained via a balanced factorization of the appropriately constructed Hankel matrix. The skewness of the singular values of the estimated Hankel matrix is applied for testing the order selection for the bilinear model.

Keywords: Stochastic system identification, Nonlinear system identification, Estimation and filtering, Realization theory, Frequency domain identification, Discrete bilinear systems, Singular values of Hankel matrix, Hermite series expansion, Order estimation by skewness.

1. INTRODUCTION

From the early seventies up to now deterministic and stochastic bilinear systems and time series are subject of intensive research, since those are the first stage beyond linearity towards nonlinearity. Bilinear systems arise as natural models for a variety of physical and biomedical chemical processes, Lakshminarayanan et al. (2001). Realization and identification of bilinear systems have applications in many fields and in particular identification of the nonlinear dynamics of an autonomous vehicle Rödönyi et al. (2021), and certain aircraft dynamics, biological systems Mohler and Kolodziej (1980), Isidori (1995) etc. Reachable and observable factorization of the Hankel matrix and minimal realization of a continuous bilinear system are due to Alessandro et al. (1974) and a similar treatment by Isidori (1973) for discrete time case. The singular value decomposition of the Hankel matrix is known to be robust therefore the balanced realization has particular importance and it is also appropriate for model reduction beside the identification, see Hsu et al. (1983); Hsu (1985), Al-Baiyat (2004), and Zhang et al. (2003). Linear and bilinear stochastic realization problem has been considered in Desai (1986), where the crosscovariances between output and stochastic input are used for realization. Wiener theory of homogeneous chaos is applied for continuous nonlinear stochastic realization problem by Lindquist et al. (1982), and in particular for a simple bilinear system which has only two terms from the infinite series. The subspace approach for identification of time-invariant multi-input discrete bilinear system with observed white noise input processes was used successfully by Favoreel et al. (1999) where the rank of the system has been established by singular values empirically, see also Verdult et al. (1998).

⋆ The research was supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Autonomous Systems National Laboratory Program.

Identification of MIMO bilinear systems driven by white noise inputs has been considered in dos Santos et al. (2009) using the fact that the bilinear term is a second-order white noise process. The cross-cumulants up to third order between the output and input for identification is applied by Tsoulkas et al. (2001). Recently Petreczky and Vidal (2018) considered the realization of bilinear systems with observed input without considering the problem of estimation of the rank of the system.

In this paper we consider the multiple Wiener–Itô representation of the stochastic stationary bilinear model driven by Gaussian white noise input, Terdik (1999). The stochastic bilinear state space has infinitely many transfer functions fulfilling a recursion in terms of system parameters, Terdik and Bokor (2010). Based on the Hermite polynomial expansion of the observation we estimate the coefficients by the help of cross-covariances between the output and the Hermite polynomials of the input. We build the Hankel matrix up of the estimated coefficients applying Isidori's (Isidori (1973)) construction, which has been worked out for deterministic case. We assume that independent samples of the input-output observations are available. Each sample provides an estimate of the Hankel matrix from which we derive the singular values. In this way we have a sample for singular values of the system. The main result of this paper is testing the rank of the estimated Hankel matrix. The test is based on the examination of the skewness of the estimated singular values of the Hankel matrix. The Appendix with the construction of the Hankel matrix in terms of the coefficients of the Hermite expansion closes the paper.

2. BILINEAR STATE SPACE MODEL

An observation Y_t is called stochastic bilinear system if it fulfils the state space model

10.1016/j.ifacol.2023.10.076

$$\mathbf{X}_t = \mathbf{A}\left(\mathbf{X}_{t-1}+\mathbf{a}\mathrm{w}_{t-1}\right)+\mathbf{D}\left(\mathbf{X}_{t-1}+\mathbf{a}\mathrm{w}_{t-1}\right)\mathrm{v}_{t-1}+\mathbf{b}\mathrm{v}_{t-1},$$
$$Y_t = \mathbf{c}^\intercal\left(\mathbf{X}_t+\mathbf{a}\mathrm{w}_t\right)+\mathrm{v}_t,$$

where $\mathbf{X}_t \in \mathbb{R}^d$, $\mathbf{A}, \mathbf{D} \in \mathbb{R}^{d\times d}$, $\mathbf{b}, \mathbf{c}, \mathbf{a} \in \mathbb{R}^d$. Notice that the state process $\mathbf{X}_t$, $t \in \mathbb{Z}$ is bilinear (Petreczky and Vidal (2018), Favoreel et al. (1999), see also Cox et al. (2018) for Markov-parameters of LPV-SS models). We assume that the output Y_t together with the scalar Gaussian white noise input v_t are observed, $\mathsf{E}\mathrm{v}_t = 0$, $\mathsf{E}\mathrm{v}_t^2 = \sigma_\mathrm{v}^2$, and the unobserved Gaussian white noise series w_t ($\mathsf{E}w_t = 0$, $\mathsf{E}w_t^2 = \sigma_w^2$) is independent of v_t.

The method of identification in this paper is based on covariances between the observation Y_t and higher order Hermite polynomials of v_t. These covariances does not contain any information on w_t. Hence we set the coefficient $\mathbf{a}$ of the noise gain to zero and postpone its estimation after the bilinear parameters have been estimated. The parameter $\mathbf{a}$ can be estimated by solving ARE-like equations afterwards. In this way we can simplify the bilinear state-space equation

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1}+\mathbf{D}\mathbf{X}_{t-1}\mathrm{v}_{t-1}+\mathbf{b}\mathrm{v}_{t-1}, \qquad (1)$$
$$Y_t = \mathbf{c}^\intercal\mathbf{X}_t+\mathrm{v}_t.$$

We assume that Y_t, $t \in \mathbb{Z}$ is physically realizable with respect to the input $\{\mathrm{v}_t, t \in \mathbb{Z}\}$. It is seen that $\mathbf{X}_t$ is independent of v_t and $\mathsf{E}\mathbf{X}_t = 0$. The most common assumption here is that all the eigenvalues of the matrix $\mathbf{A}\otimes\mathbf{A}+\sigma_\mathrm{v}^2\mathbf{D}\otimes\mathbf{D}$ are less than one in modulus (where $\otimes$ denotes the Kronecker matrix product). This is the necessary and sufficient condition for the existence of a second order stationary physically realizable solution of (1) as well, see Terdik (1985), Liu and Brockwell (1988) and Terdik (1999). The infinite Hankel matrix $\mathbf{H}$ of the transfer function system $\{\mathbf{c}^\intercal\mathbf{g}_k\}$, defined below, has finite rank if and only if there exits bilinear realization. The rank is the dimension of the minimal realization and an actual minimal realization is provided by the quadruplet $\mathbf{A}, \mathbf{D}, \mathbf{b}, \mathbf{c}$, see Isidori (1973), Isidori (1995) Theorem 3.4.3, p. 127. In frequency domain a large class of stationary stochastic series $\mathbf{X}_t$, which belongs to the nonlinear space generated by the Gaussian white noise series v_t, has Wiener–Itô representation

$$\mathbf{X}_t = \sum_{r=1}^{\infty}\int_{\mathcal{D}^r} e^{i2\pi t\sum_{j=1}^{r} f_j}\mathbf{g}_r(f_{1:r})W\left(df_{1:r}\right), \qquad t \in \mathbb{Z}, (2)$$
$$Y_t = \mathbf{c}^\intercal\mathbf{X}_t+\mathrm{v}_t.$$

where the integral of a vector valued function is meant by coordinate wise, $\mathcal{D} = [0,1]$, $f_{1:r} = (f_1,\ldots,f_r)$, $\mathbf{g}_r$ denote the transfer functions of $\mathbf{X}_t$ and $W\left(df_{1:r}\right)$ is the multiple stochastic spectral measure with respect to the Gaussian white noise series $\{\mathrm{v}_t, t \in \mathbb{Z}\}$. We recall that multiple Wiener–Itô integrals in the representation (2) have the following properties (see Terdik (1999), Sect. 2 for details)

(1) Wiener–Itô integrals are real valued and $\mathsf{E}\int_{\mathcal{D}^r} g_r(f_{1:r})W\left(df_{1:r}\right) = 0$,

(2) Wiener–Itô integrals with different orders are orthogonal: if $r \neq q$, then

$$\mathsf{E}\left(\int_{\mathcal{D}^r} g_r(f_{1:r})W\left(df_{1:r}\right)\int_{\mathcal{D}^q} g_q(f_{1:q})W\left(df_{1:q}\right)\right) = 0,$$

and

$$\operatorname{Var}\left(\int_{\mathcal{D}^r} g_r(f_{1:r})W\left(df_{1:r}\right)\right) = \sigma_\mathrm{v}^{2r} r!\int_{\mathcal{D}^r}\left|\operatorname{sym} g_r\left(f_{1:r}\right)\right|^2\prod_{k=1}^{r} s\left(f_k\right)df_{1:r},$$

(3) If $t_1, t_2, \ldots, t_r \in \mathbb{R}$

$$\int_{\mathcal{D}^r} e^{i2\pi\sum_{j=1}^{r} t_j f_j} W\left(df_{1:r}\right) = H_r\left(\mathrm{v}_{t_1}, \mathrm{v}_{t_2}, \ldots, \mathrm{v}_{t_r}\right),$$

where $H_r\left(\mathrm{v}_{t_1}, \mathrm{v}_{t_2}, \ldots, \mathrm{v}_{t_r}\right)$ denotes the r^{th} degree Hermite polyinomial of Gaussian variables $\mathrm{v}_{t_1},\ldots,\mathrm{v}_{t_r}$.

There will be no confusion if we denote both $e^{-i2\pi f}$ and the back shift operator by z^{-1}, in this sense $\mathsf{z}_1^{-k_1}\mathsf{z}_2^{-(k_1+k_2)} = e^{-i2\pi(k_1 f_1+(k_1+k_2)f_2)}$ and $\mathsf{z}_{1:2}^{-(k_1,k_1+k_2)}\mathrm{v}_t = \left[\mathrm{v}_{t-k_1}, \mathrm{v}_{t-(k_1+k_2)}\right]$.

The first transfer function $\mathbf{g}_1(f_1)$ in (2) corresponds to the linear part of the state variate $\mathbf{X}_t$ in (1), the second one contains the contribution of all possible second order products of the input and so on. The following recursive formula for the transfer functions can be derived easily:

$$\mathbf{g}_1(f_1) = \left(\mathsf{z}_1\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{b},$$
$$\mathbf{g}_2(f_{1:2}) = \left(\mathsf{z}_1\mathsf{z}_2\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{D}\mathbf{g}_1(f_1),$$

and in general for $r \geq 2$,

$$\mathbf{g}_r\left(f_{1:r}\right) = \left(\mathsf{z}_1\cdots\mathsf{z}_r\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{D}\mathbf{g}_{r-1}\left(f_{1:r-1}\right).$$

The first transfer function is clearly the linear part of the system and corresponds to the series

$$\int_{\mathcal{D}} e^{i2\pi t f_1}\mathbf{g}_1\left(f_1\right)W\left(df_1\right) = \int_{\mathcal{D}} e^{i2\pi t f_1}\left(\mathsf{z}_1\mathbf{I}-\mathbf{A}\right)^{-1}W\left(df_1\right)\mathbf{b} = \sum_{k=0}^{\infty}\mathbf{A}^k\int_{\mathcal{D}}\mathsf{z}_1^{t-(k+1)}W\left(df_1\right)\mathbf{b} = \sum_{k=0}^{\infty}\mathbf{A}^k\mathrm{v}_{t-(k+1)}\mathbf{b}.$$

The linear part of $\mathbf{X}_t$ is written in terms of first order Hermite polynomial is $H_1\left(\mathrm{v}_{t-(k+1)}\right) = \mathrm{v}_{t-(k+1)}$ as

$$\sum_{k=0}^{\infty}\mathbf{A}^k\mathbf{b}H_1\left(\mathrm{v}_{t-(k+1)}\right) = \mathbf{b}\mathrm{v}_{t-1}+\mathbf{A}\mathbf{b}\mathrm{v}_{t-2}+\ldots.$$

Consider the series expansion of Y_t then the coefficient of v_t is 1, the coefficient of v_{t-1} is $\mathbf{c}^\intercal\mathbf{b}$ let it be denoted by ℓ_0. In general the coefficients $\mathbf{c}^\intercal\mathbf{A}^k\mathbf{b}$ of the linear part of the series expansion of Y_t will be denoted by $h\left(\mathbf{0}_k\right)$, where $\mathbf{0}_k$ denotes k consecutive zeros. In particular if $k = 1$ we set $h(0) = \mathbf{c}^\intercal\mathbf{A}\mathbf{b}$, and so on. The second order transfer function

$$\mathbf{g}_2\left(f_{1:2}\right) = \left(\mathsf{z}_1\mathsf{z}_2\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{D}\left(\mathsf{z}_1\mathbf{I}-\mathbf{A}\right)^{-1}\mathbf{b},$$

has the series expansion

$$\mathbf{g}_2\left(f_{1:2}\right) = \sum_{k_{1:2}=0}^{\infty}\mathbf{A}^{k_2}\mathbf{D}\mathbf{A}^{k_1}\mathbf{b}\mathsf{z}_1^{-(k_1+k_2+2)}\mathsf{z}_2^{-(k_1+1)}.$$

Let the coefficient $\mathbf{c}^\intercal\mathbf{A}^{k_2}\mathbf{D}\mathbf{A}^{k_1}\mathbf{b}$ of the Hermite polynomial $H_2\left(\mathrm{v}_{t-(k_1+k_2+2)}, \mathrm{v}_{t-(k_1+1)}\right) = H_2\left(\mathsf{z}_{1:2}^{-(k_1+k_2+2,k_1+1)}\mathrm{v}_t\right) = \mathrm{v}_{t-(k_1+k_2+2)}\mathrm{v}_{t-(k_1+1)}$, be denoted by $h\left(\mathbf{0}_{k_1}1\mathbf{0}_{k_2}\right)$. Notice that exponents of $\mathbf{A}$ are indexed by opposite (right to left) order. An example is that if $k_{1:2} = 0$ then the coefficient of $H_2(\mathrm{v}_{t-1}, \mathrm{v}_{t-2})$ is $h(1) = \mathbf{c}^\intercal \mathbf{D}\mathbf{b}$.

Now set

$$\mathbf{g}_3(f_{1:3}) = (\mathsf{z}_1\mathsf{z}_2\mathsf{z}_3\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}\mathbf{g}_2(f_{1:2})$$
$$= \sum_{k_{1:3}=0}^{\infty} \mathbf{A}^{k_3}\mathbf{D}\mathbf{A}^{k_2}\mathbf{D}\mathbf{A}^{k_1}\mathbf{b}\mathsf{z}_{1:3}^{-(k_1+k_2+k_3+3,k_2+k_3+2,k_3+1)}.$$

We denote the coefficient $\mathbf{c}^\intercal\mathbf{A}^{k_3}\mathbf{D}\mathbf{A}^{k_2}\mathbf{D}\mathbf{A}^{k_1}\mathbf{b}$ of Y_t by $h(\mathbf{0}_{k_1}1\mathbf{0}_{k_2}1\mathbf{0}_{k_3})$. In particular $h(\mathbf{1}_2) = \mathbf{c}^\intercal\mathbf{D}^2\mathbf{b}$ $(k_{1:3} = 0)$ corresponds to the $H_3(\mathrm{v}_{t-1}, \mathrm{v}_{t-2}, \mathrm{v}_{t-3})$, where $\mathbf{1}_k$ denotes k consecutive 1s. Notice that the term $\mathsf{z}_1^{-(k_1+k_2+k_3+3)}\mathsf{z}_2^{-(k_2+k_3+2)}\mathsf{z}_3^{-(k_3+1)}$ in series expansion of $\mathbf{g}_3$ guarantees that all the indices of the variables of H_3 are distinct. It is also seen that the degree of the Hermite polynomial is the number of ones plus one, moreover the Hermite polynomials are symmetric therefore it will be more convenient renumber the exponents, say instead of $\mathsf{z}_1^{-(k_1+k_2+k_3+3)}\mathsf{z}_2^{-(k_2+k_3+2)}\mathsf{z}_3^{-(k_3+1)}$ we shall put $\mathsf{z}_1^{-(k_1+1)}\mathsf{z}_2^{-(k_1+k_2+2)}\mathsf{z}_3^{-(k_1+k_2+k_3+3)}$.

In general we introduce a set $(\mathbf{0}_{k_1}1\mathbf{0}_{k_2}1\mathbf{0}_{k_3}\ldots 1\mathbf{0}_{k_K})$ for indices of Hermite polynomials where there are exactly $K-1 \geqq 0$ ones between consecutive zeros $\mathbf{0}_{k_j}$, see in the Appendix as well. Some blocs of zeros can be empty $(k_j = 0)$. If $K = 1$, then we have $(\mathbf{0}_k)$. If $K > 0$, and $k_{1:K} = 0$, then we have the set $(\mathbf{1}_{K-1})$. The series expansion of the transfer function $\mathbf{c}^\intercal\mathbf{g}_K(f_{1:K})$ can be described by the help of these indices. Namely the coefficient of Hermite polynomial $H_K\left(\mathsf{z}_{1:K}^{-(k_1+1,k_1+k_2+2,\ldots\Sigma k_\ell+K)}\mathrm{v}_t\right)$ is $\mathbf{c}^\intercal\mathbf{A}^{k_K}\mathbf{D}\ldots\mathbf{A}^{k_2}\mathbf{D}\mathbf{A}^{k_1}\mathbf{b}$ which corresponds to the index $(\mathbf{0}_{k_1}1\mathbf{0}_{k_2}1\mathbf{0}_{k_3}\ldots 1\mathbf{0}_{k_K})$ and will be denoted by $h(\mathbf{0}_{k_1}1\mathbf{0}_{k_2}1\mathbf{0}_{k_3}\ldots 1\mathbf{0}_{k_K})$. In particular $h(\mathbf{1}_K) = \mathbf{c}^\intercal\mathbf{D}^{K-1}\mathbf{b}$ is the coefficient of $H_K(\mathrm{v}_{t-1}, \mathrm{v}_{t-2}, \ldots, \mathrm{v}_{t-K})$ in the series expansion of $\mathbf{X}_t$. Note that these indices can be considered as digits of integers. Now we can build up the Hankel matrix $\mathbf{H}$ of the system using the coefficients $h(\mathbf{0}_{k_1}1\mathbf{0}_{k_2}1\mathbf{0}_{k_3}\ldots 1\mathbf{0}_{k_K})$, Isidori (1973). A stationary stochastic series Y_t which has Wiener–Itô representation (2) with respect to a Gaussian white noise series v_t can be associated with a Hankel matrix with entries $\mathrm{Cov}\left(Y_t, H_K\left(\mathbf{z}^{-_cS(k_{1:K}+\mathbf{1})}\mathrm{v}_t\right)\right)$. Then one can truncate the singular values of the Hankel matrix for getting a bilinear approximation of the process Y_t.

3. ESTIMATION OF THE HANKEL MATRIX

Recall that we have assumed that the input series v_t is *Gaussian white noise* with mean 0 and variance σ_v^2. One can get this assumption fulfilled with prewithening the input Gaussian process. The estimation of the entries $h(i_{1:k})$ of the Hankel matrix follows from the cross covariances between the output and the Hermite polynomials of the input v_t. We introduce the following short notation: let the cumulative sum $\mathrm{cumSum}(k_1+1, k_2+1, k_3+1, \ldots, k_K+1)$ be denoted by $_cS(k_{1:K}+\mathbf{1})$, and $\left(\mathrm{v}_{t-(k_1+1)}, \mathrm{v}_{t-(k_1+k_2+2)}, \ldots \mathrm{v}_{t-(\Sigma k_j+K)}\right)$ $= \mathbf{z}^{-_cS(k_{1:K}+\mathbf{1})}\mathrm{v}_t$. An example is $K = 3$, when $_cS(k_{1:3}+\mathbf{1}_3) = \mathrm{cumSum}(k_1+1, k_2+1, k_3+1)$ $= (k_1+1, k_1+k_2+2, k_1+k_2+k_3+3)$. It is seen that each $k_j \geq 0$, nevertheless all indices of the input v_t are distinct, strictly decreasing, and it starts from $t-(k_1+1)$. We have the corresponding Hermite polynomials

$$H_K\left(\mathsf{z}_{1:K}^{-_cS(k_{1:K}+\mathbf{1}_K)}\mathrm{v}_t\right) = H_K\left(\mathrm{v}_{t-(k_1+1)}, \mathrm{v}_{t-(k_1+k_2+2)}, \ldots\right).$$

These Hermite polynomials constitute an orthogonal system therefore we can apply the equation

$$\mathrm{Cov}\left(Y_t, H_K\left(\mathbf{z}^{-_cS(k_{1:K}+\mathbf{1})}\mathrm{v}_t\right)\right)$$
$$= \mathrm{Cov}\left((\mathbf{c}^\intercal\mathbf{X}_t + \mathrm{v}_t), H_K\left(\mathbf{z}^{-_cS(k_{1:K}+\mathbf{1})}\mathrm{v}_t\right)\right)$$
$$= \sigma_\mathrm{v}^{2K}\mathbf{c}^\intercal\mathbf{A}^{k_K}\mathbf{D}\ldots\mathbf{A}^{k_2}\mathbf{D}\mathbf{A}^{k_1}\mathbf{b},$$

for estimating the entries of the Hankel matrix. The variance σ_v^2 can be estimated from the input. As we have seen the coefficients $h(\cdot)$ of the Hermite polynomials in the series expansion (1) of the input process Y_t provides the Hankel matrix.

4. TEST FOR THE HANKEL RANK

The singular values $\mathfrak{s}$ of the Hankel matrix $\mathbf{H}$ give a possible estimation for the Hankel rank of the system. The singular values are non-negative and the default orders is monotone decreasing. We assume that either independent samples of the input-output observations are available or we have that the number of observations is so large that it can be sliced such that each slice is large enough for providing an estimate of $\mathfrak{s}$. We estimate the Hankel matrix $\mathbf{H}$ and its singular values $\mathfrak{s}^{(L)}$ for each sample. In this way we have an independent sample $\widehat{\mathfrak{s}}_k^{(L)}$ for each singular value $\mathfrak{s}^{(L)}$. This sample follows the CLT and asymptotically $\overline{\widehat{\mathfrak{s}}_k^{(L)}}$ is approaching the normal distribution. In case the expected value of $\widehat{\mathfrak{s}}_k^{(L)}$ is zero the sample of a normal distribution would take a symmetrical values with respect to zero. Which can not happen for a sample of zero singular value since it should be non-negative. This implies that the closer to zero a singular value is, the more its distribution skewed. The skewness of a distribution is well studied, the definition is the 3^{rd} order cumulant of the standardized variates. This is the first characteristic of discrepancy form normal variate since all higher (then 2) order cumulants are zero for a normal variate. Let us suppose that we have a sample $\widehat{\mathfrak{s}}_k^{(L)}$, $k = 1:N$ for the L^{th} singular value $\mathfrak{s}^{(L)}$ of the Hankel matrix $\mathbf{H}$. Our empirical findings is showing clearly the fact that as far as a singular value $\mathfrak{s} = 0$ the distribution of the estimated singular value $\widehat{\mathfrak{s}}$ skewed to zero compared to the cases when $\mathfrak{s} \neq 0$, see Figure 1. We estimate the index of skewness $\gamma_{1,L}$ of the L^{th} singular value by

$$\widehat{\gamma}_{1,L} = \overline{\left(\frac{\widehat{\mathfrak{s}}_k^{(L)} - \overline{\widehat{\mathfrak{s}}_k^{(L)}}}{\sigma_N\left(\widehat{\mathfrak{s}}_k^{(L)}\right)}\right)^3},$$

where $\sigma_N\left(\widehat{\mathfrak{s}}_k^{(L)}\right)$ denotes the sample standard deviation of the sample $\widehat{\mathfrak{s}}_k^{(L)}$. The most popular test for checking the skewness $\gamma_1 = 0$ has been given by Mardia (1970): using asymptotic normality of $\widehat{\gamma}_{1,L}$ the $N\widehat{\gamma}_{1,L}^2/6$ proves to be chi square distributed with degree of freedom 1. The asymptotic variance of the estimated skewness $\widehat{\gamma}_{1,L}$ of $\widehat{\mathfrak{s}}_k^{(L)}$ is also known to be

$$\mathrm{var}(\widehat{\gamma}_{1,L}) = cum_{\mathfrak{s},6} + 9cum_{\mathfrak{s},4} + 9cum_{\mathfrak{s},3}^2 + 6,$$

where $cum_{\mathfrak{s},k}$ are the k^{th} order cumulants of $\widehat{\mathfrak{s}}_k^{(L)}$ (see Terdik (2021), p. 345).

Now we perform the test $H_{0,L}$: $\mathfrak{s}^{(L)} \neq 0$, against $H_{1,L}$: $\mathfrak{s}^{(L)} = 0$, i.e. the singular value $\mathfrak{s}^{(L)}$ is non zero, i.e. $\widehat{\mathfrak{s}}^{(L)}$ is unskewed. Hence we test the skewness of $\widehat{\mathfrak{s}}_L$ with the test statistic $\widehat{\gamma}_{1,L}$, such that $H_{0,L}$ is rejected if the skewness $\gamma_{1,L}$ of the distribution of $\widehat{\mathfrak{s}}_k^{(L)}$ is not zero. Under the $H_{0,L}$ hypothesis the variance is reduced to $cum_{\mathfrak{s},6} + 9cum_{\mathfrak{s},4} + 6$ and it is estimated by the sample variance of Hermite polynomial $H_3\left(\widehat{\mathfrak{s}}_k^{(L)}\right)$. If $\mathfrak{s}^{(L)} \neq 0$ then $\widehat{\gamma}_{1,L}^2 / \operatorname{var}\left(\widehat{\gamma}_{1,L}\right)$ is asymptotically chi square distributed with degree of freedom 1, (see Terdik (2021), (6.22)). Note that we use this test recursively by L, $L-1$, ... until the $\mathfrak{s}^{(L)} \neq 0$ hypothesis is accepted. The Figure 1 shows that decreasing L the histogram becomes unskewed and the $p-$values are increasing. The result is that we stop at $L = 3$ and except that the order of the system is 3. The method of truncating the singular values of the Hankel matrix $\mathbf{H}$ with entries $EY_t H_K\left(\mathbf{z}^{-_c S(k_{1:K}+\mathbf{1})}\mathrm{v}_t\right)$ provides bilinear approximation for a nonlinear process Y_t.

5. SIMULATION

In this section we consider the bilinear system with state space representation form

$$\begin{aligned} \mathbf{X}_t &= A\mathbf{X}_{t-1} + D\mathbf{X}_{t-1}\mathrm{v}_{t-1} + \mathbf{b}\mathrm{v}_{t-1}, \\ Y_t &= \mathbf{c}'\mathbf{X}_t. \end{aligned} \tag{3}$$

We put $\mathfrak{s}_{\mathrm{v}}^2 = 2$, $\mathbf{A} = \begin{bmatrix} 0.15 & 1 & 0 \\ 0.1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, the bilinear parameters $\mathbf{D} = \begin{bmatrix} 0.3 & 0 & -0.2 \\ 0.5 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, finally $\mathbf{b} = [0.6, 0, 0]^\top$ and $\mathbf{c} = [1, 0, 0]^\top$. We calculated the Hankel matrix $\mathbf{H}$ with these parameters. The singular values $\mathfrak{s}$ of $\mathbf{H}$ are $[1.0067, 0.4612, 0.1761]$ except zeros. Using these singular values $\mathbf{Q}_6$ and $\mathbf{P}_6$ are calculated, such that both Gramians are the singular values $\mathfrak{s}$. The new parameters proved to be $\widetilde{\mathbf{b}} = [-0.7509, 0.2393, -0.0622]^\top$, $\widetilde{\mathbf{c}} = [-0.76060.14910.1101]^\top$,

$$\widetilde{\mathbf{A}} = \begin{bmatrix} 0.3214 & 0.6292 & 0.0228 \\ 0.0692 & -0.2341 & -0.0321 \\ -0.3939 & 0.1346 & 0.0628 \end{bmatrix},$$

and $\widetilde{\mathbf{D}} = \begin{bmatrix} 0.3814 & -0.1223 & 0.3750 \\ 0.6045 & -0.1034 & -0.2246 \\ 0.0948 & -0.0225 & 0.0219 \end{bmatrix}$. Note that these new parameters provide the same Hankel matrix.

Now we put the sample size as 2^{10} and repeated the estimations 2^7 times. The Hankel matrix with $L = 6$ is estimated using the cross-covariances. We found the estimated singular values $\widehat{\mathfrak{s}} = [1.0036, 0.4924, 0.1937, 0.1051, 0.0900, 0.0779]$ see the corresponding p-values by the Figure 1. Hence we except that the rank of $\widehat{\mathbf{H}}$ is 3. The estimated system parameters as follows:

$\widehat{\mathbf{b}} = [-0.7433, 0.2525, -0.0655]$,

$\widehat{\mathbf{c}} = [-0.7575, 0.1274, 0.1155]$,

$$\widehat{\mathbf{A}} = \begin{bmatrix} 0.3070 & 0.6511 & 0.0463 \\ 0.1191 & -0.2231 & -0.0681 \\ -0.4049 & 0.1028 & 0.0198 \end{bmatrix},$$

and

$$\widehat{\mathbf{D}} = \begin{bmatrix} 0.3705 & -0.1572 & 0.3889 \\ 0.6130 & -0.0864 & -0.2166 \\ 0.1237 & -0.0211 & -0.0897 \end{bmatrix},$$

compare to the true values $\widetilde{\mathbf{b}}, \widetilde{\mathbf{c}}, \widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}$.

6. CONCLUSIONS

This paper proposed a realization - based identification procedure for stochastic bilinear systems. The Hankel matrix of the system is built up in terms of the Hermite series representation of the process. The estimation of the entries of the Hankel matrix is obtained from the cross covariances between the output and the Hermite polynomials of the input. A statistical test is performed which based on the skewness of the distribution of the estimated Hankel singular values. This idea is elaborated to get the order of the bilinear model.

In practice one slices the observation and estimate the singular values for each slice and use this sample of singular values to test the Hankel rank of the system. If the Hankel rank is given then the parameters $\mathbf{A}$, $\mathbf{D}$, $\mathbf{b}$, and $\mathbf{c}$ are calculated from the estimated Hankel matrix resulting in a balanced realization. Based on the Hankel matrix for the general nonlinear model (2) one can use the method of Section 3 for getting either an approximate bilinear model or a model reduction of a more general nonlinear model to the bilinear.

REFERENCES

Al-Baiyat, S.A. (2004). Model reduction of bilinear systems described by input–output difference equation. *International Journal of Systems Science*, 35(9), 503–510.

Alessandro, D.P., Isidori, A., and Ruberti, A. (1974). Realization and structure theory of bilinear dynamical systems. *SIAM J. Control*, 12(3), 517–535.

Cox, P.B., Tóth, R., and Petreczky, M. (2018). Towards efficient maximum likelihood estimation of LPV-SS models. *Automatica*, 97, 392–403. doi: 10.1016/j.automatica.2018.08.021.

Desai, U. (1986). Realization of bilinear stochastic systems. *IEEE Trans. Aut. Contr.*, AC-31, 189–192.

dos Santos, P., Ramos, J., and de Carvalho, J. (2009). Identification of bilinear systems with white noise inputs: An iterative deterministic-stochastic subspace approach. *IEEE Transactions on Control Systems Technology*, 17(5), 1145–1153. doi: 10.1109/tcst.2008.2002041.

Favoreel, W., De Moor, B., and Van Overschee, P. (1999). Subspace identification of bilinear systems subject to white inputs. *IEEE Transactions on Automatic Control*, 44(6), 1157–1165.

Hsu, C., Desai, U., and Crawley, C. (1983). Realization algorithms and approximation methods of bilinear systems. In *The 22nd IEEE Conference on Decision and Control*. IEEE. doi:10.1109/cdc.1983.269628.

Hsu, C. (1985). Realization and approximation of discrete bilinear systems. *Applied digital control*, 171–187.

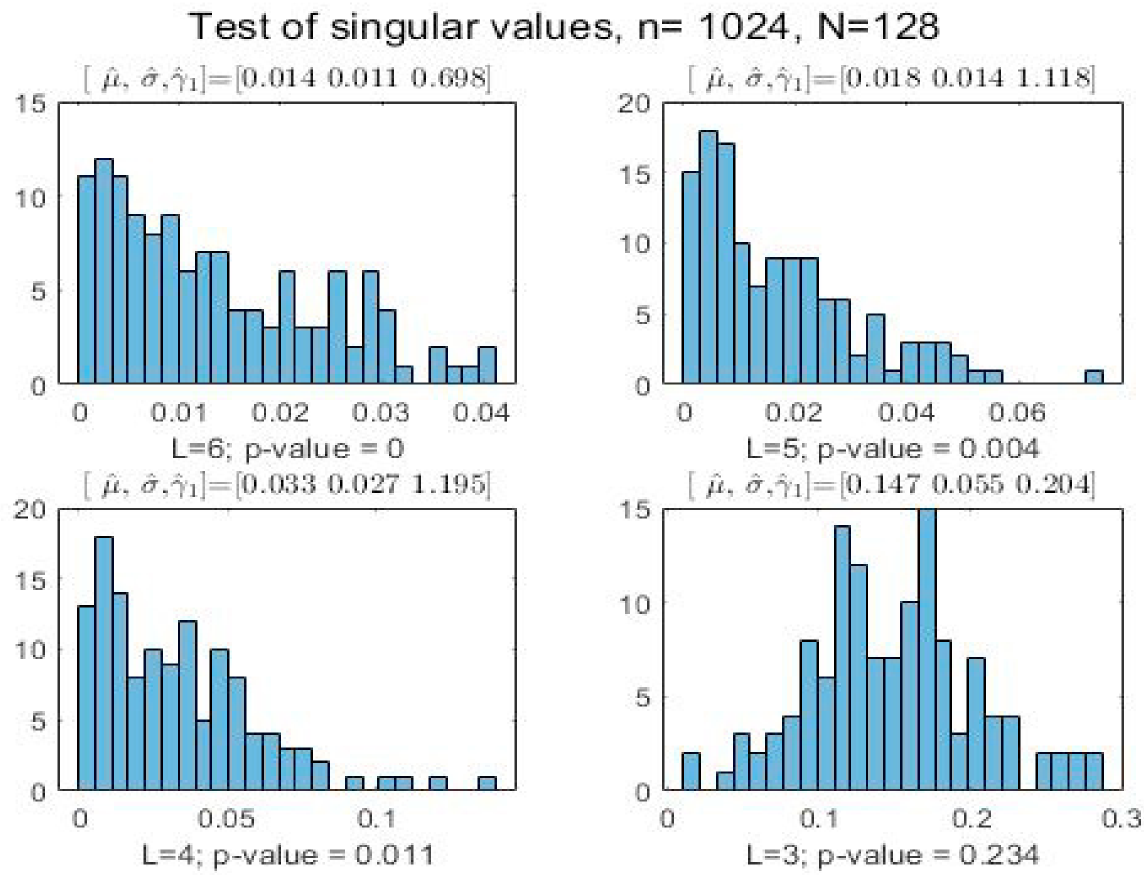

Fig. 1. Histograms for the model in Section A, where the x-axis denotes the real line and y-axis shows the values of the histogram.

Isidori, A. (1973). Direct construction of minimal bilinear realizations from nonlinear input output maps. *IEEE Trans. Aut. Contr.*, AC-18, 626–631.

Isidori, A. (1995). *Nonlinear control systems.* Communications and Control Engineering Series. Springer-Verlag, Berlin, third edition.

Lakshminarayanan, S., Mhatre, P., and Gudi, R. (2001). Identification of bilinear models for chemical processes using canonical variate analysis. *Industrial & Engineering Chemistry Research*, 40(20), 4415–4427. doi: 10.1021/ie000685b.

Lindquist, A., Mitter, S., and Picci, G. (1982). Toward a theory of nonlinear stochastic realization. In *Feedback Control of Linear and Nonlinear Systems*, 175–189. Springer.

Liu, J. and Brockwell, P.J. (1988). On the general bilinear time series model. *J. Appl. Prob.*, 25, 553–564.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530.

Mohler, R.R. and Kolodziej, W.J. (1980). An overview of bilinear system theory and applications. *IEEE Trans. Systems Man. Cybern.*, SMC-10, 683–688.

Petreczky, M. and Vidal, R. (2018). Realization theory for a class of stochastic bilinear systems. *IEEE Transactions on Automatic Control*, 63(1), 69–84. doi: 10.1109/tac.2017.2710801.

Rödönyi, G., Beintema, G., Tóth, R., Schoukens, M., Pup, D., Kisari, Á., Vígh, Z., Körös, P., Soumelidis, A., and Bokor, J. (2021). Identification of the nonlinear steering dynamics of an autonomous vehicle. *IFAC-PapersOnLine*, 54(7), 708–713. doi: 10.1016/j.ifacol.2021.08.444.

Terdik, Gy. (1985). Transfer functions and conditions for stationarity of bilinear models with gaussian white noise. *Proc. R. Soc. London A*, 400, 315–330.

Terdik, Gy. (1999). *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis; A Frequency Domain Approach*, volume 142 of *Lecture Notes in Statistics*. Springer Verlag, New York.

Terdik, Gy. (2021). *Multivariate Statistical Methods; Going Beyond the Linear*. Frontiers in Probability and the Statistical Sciences. Springer International Publishing, Cham, Switzerland. doi:10.1007/978-3-030-81392-5.

Terdik, Gy. and Bokor, J. (2010). Stochastic input-output realization of bilinear systems. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS*, volume 5, 1–6.

Tsoulkas, V., Koukoulas, P., and Kalouptsidis, N. (2001). Identification of input-output bilinear systems using cumulants. *IEEE Transactions on Signal Processing*, 49(11), 2753–2761.

Verdult, V., Verhaegan, M.H., Chou, C.T., and Lovera, M. (1998). Efficient and systematic identification of MIMO bilinear state space models. In *Proceedings of the 37th IEEE Conference on Decision and Control*, Cat. No.98CH36171, 1260–1265. IEEE. doi:

10.1109/cdc.1998.758451.
Zhang, L., Lam, J., Huang, B., and Yang, G.H. (2003). On gramians and balanced truncation of discrete-time bilinear systems. *International Journal of Control*, 76(4), 414–427. doi:10.1080/0020717031000082540.

Appendix A. HANKEL MATRIX FOR BILINEAR SYSTEMS AND BALANCED CANONICAL FORM

The entries $h(\cdot)$ of the Hankel matrix $\mathbf{H}$ will be indexed by vectors of 0s and 1s. Both the first row and the first column will be generated by the similar recursion. Let the first entry be the empty vector $\emptyset$. The next 2 entries are the vectors (0) and (1) then the next 4 entries are the binary digits of numbers $0, 1, \ldots, 3$, namely $(\mathbf{0}_2), (01), (10), (\mathbf{1}_2)$, where $\mathbf{0}_k$ and $\mathbf{1}_k$ denote k consecutive 0s and 1s and so on. In general the k^{th} blocks of indices are the binary digit vectors of numbers $0, 1, \ldots, 2^k - 1$, for $k = 1, \ldots$. We allow arbitrary orders of numbers within a block. Finally the index vector of an entry of $\mathbf{H}$ is the concatenation of the corresponding index vectors of the first column and the first row respectively. Notice that this is true for the first row and the first column as well since the first entry in both cases is an empty vector.

Now we show an example of the Hankel matrix which is used in this paper.

Example 1. The first block of the first column is (0) and (1) then we generate each block adding an extra 0 to the front of each index vector of the previous block then repeat these steps with 1 instead of 0. The first row starts with (1) and (0) then follow the construction of the blocks of the first row with 1 first then 0. Let the dimension d, matrices $\mathbf{A}$, $\mathbf{D} \in \mathbb{R}^{d\times d}$, and vectors $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^{d\times d}$ be given. Then we can build up the Hankel matrix $\mathbf{H}$ in the following way. Let us identify 0 with $\mathbf{A}$ and 1 with $\mathbf{D}$ then let $h(\mathbf{0}_{k_1}\mathbf{1}_{m_1}\ldots\mathbf{0}_{k_K}\mathbf{1}_{m_M}) = \mathbf{c}^\intercal\mathbf{A}^{k_1}\mathbf{D}^{m_1} \times \cdots {}^\intercal\mathbf{A}^{k_K}\mathbf{D}^{m_M}\mathbf{b}$. The construction of the Hankel matrix $\mathbf{H}$ guarantee the existence of matrices $\mathbf{Q}$ and $\mathbf{P}$ such that $\mathbf{H} = \mathbf{Q}\times\mathbf{P}$. It is clear that the Hankel matrix is well defined by the coefficients $h(\cdot)$ and depends on the rank of the system (d) by an intimate way, as we shall see later. The constructions of Example 1 implies that the observability matrix is $\mathbf{Q} = \left[\mathbf{c}^\intercal; \mathbf{c}^\intercal\mathbf{A}; \mathbf{c}^\intercal\mathbf{D}; \mathbf{c}^\intercal\mathbf{A}^2; \mathbf{c}^\intercal\mathbf{AD}; \mathbf{c}^\intercal\mathbf{DA}; \ldots\right]$ (';' implies to arrange the rows underneath to each other) and the reachability matrix is $\mathbf{P} = \left[\mathbf{b}, \mathbf{Db}, \mathbf{Ab}, \mathbf{D}^2\mathbf{b}, \mathbf{DAb}, \mathbf{ADb}, \ldots\right]$, see Isidori (1995). We use the following recursive algorithm, with respect to Example 1, for construction of the Hankel matrix $\mathbf{H}_{m,k} = \mathbf{Q}_m \times \mathbf{P}_k$ with dimension $\left(2^{m+1}-1\right)\times\left(2^{k+1}-1\right)$. The dimensions of $\mathbf{q}_j$ and $\mathbf{p}_j$ are $2^j \times d$, and $d \times 2^j$ hence $\mathbf{Q}_j$and $\mathbf{P}_j$ are $\left(2^{j+1}-1\right)\times d$ and $d\times\left(2^{j+1}-1\right)$ respectivelly. The factorization of $\mathbf{H}_{m,k}$ is not unique we provide a particular factorization wich results the balanced canonical form.

We apply the singular value decomposition of $\mathbf{H}_{m,k}$ with dimension $\left(2^{k+1}-1\right)\times\left(2^{m+1}-1\right)$. Define $\widetilde{\mathbf{Q}}_m$ with $\left(2^{m+1}-1\right)\times\widetilde{d}$ and $\widetilde{\mathbf{P}}_k$ with $\widetilde{d}\times\left(2^{k+1}-1\right)$ by the singular decomposition where $\widetilde{d}$ the is the number of the nonzero singular values such. We can use the square root of singular values such that the observability Gramian $\widetilde{\mathbf{Q}}_k^\intercal\widetilde{\mathbf{Q}}_k$ and

Algorithm 1 Hankel matrix construction

$\mathbf{q}_0 \leftarrow \mathbf{c}^\intercal$, $\mathbf{Q}_0 \leftarrow \mathbf{q}_0$,
$\mathbf{p}_0 \leftarrow \mathbf{b}$, $\mathbf{P}_0 \leftarrow \mathbf{p}_0$
for $j = 1 : k$ **do**
  $\mathbf{q}_j \leftarrow [\mathbf{q}_{j-1}[\mathbf{A}, \mathbf{D}]]^\intercal$;
  $\mathbf{q}_j \leftarrow$ Reshape $\mathbf{q}_j$ to $d \times 2^k$ then transpose
  $\mathbf{Q}_j \leftarrow [\mathbf{Q}_{j-1}; \mathbf{q}_j]$;
end for
for $j = 1 : m$ **do**
  $\mathbf{p}_j \leftarrow \left[\mathbf{Dp}_{j-1}, \mathbf{Ap}_{j-1}\right]$; $\mathbf{P}_j \leftarrow [\mathbf{P}_{j-1}, \mathbf{p}_j]$;
end for
$\mathbf{H}_{m,k} = \mathbf{Q}_m \times \mathbf{P}_k$.

reachability Gramian $\widetilde{\mathbf{P}}_m^\intercal\widetilde{\mathbf{P}}_m$ become diagonal and the diagonals are equal. We have $\mathbf{H}_{m,k} = \widetilde{\mathbf{Q}}_m\times\widetilde{\mathbf{P}}_k$.

Now we can select the rows and columns of $\mathbf{H}_{m,k}$ and get the matrices $\mathbf{H}_{A,k,m}$ and $\mathbf{H}_{D,k,m}$ with the property $\mathbf{H}_{A,k,m} = \widetilde{\mathbf{Q}}_{A,k}\widetilde{\mathbf{A}}\widetilde{\mathbf{P}}_m$ and $\mathbf{H}_{D,k,m} = \widetilde{\mathbf{Q}}_k\widetilde{\mathbf{D}}\widetilde{\mathbf{P}}_{D,m}$ respectively.

Algorithm 2 Selecting indices for rows of $\widetilde{\mathbf{Q}}_{A,k}$ and columns of $\widetilde{\mathbf{P}}_{D,m}$

$indexRows \leftarrow [2 : 2 : size(\mathbf{H}_{m,k}, 1)]$
$indexColumns \leftarrow [2]$
for $k = 1 : (m-2)$ **do**
  $indexColumns \leftarrow [indexColumns, 2^{k+1}, 2^{k+1} + (1 : (2^k - 1))]$;
end for

Hence we obtain the balanced canonical form
$\widetilde{\mathbf{A}} = \left(\widetilde{\mathbf{Q}}_{A,k}^\intercal\widetilde{\mathbf{Q}}_{A,k}\right)^{-1}\widetilde{\mathbf{Q}}_k^\intercal\mathbf{H}_{A,k,m}\widetilde{\mathbf{P}}_m^\intercal\left(\widetilde{\mathbf{P}}_m\widetilde{\mathbf{P}}_m^\intercal\right)^{-1}$
$\widetilde{\mathbf{D}} = \left(\widetilde{\mathbf{Q}}_k^\intercal\widetilde{\mathbf{Q}}_k\right)^{-1}\widetilde{\mathbf{Q}}_k^\intercal\mathbf{H}_{D,k,m}\widetilde{\mathbf{P}}_{D,m}^\intercal\left(\widetilde{\mathbf{P}}_{D,m}\widetilde{\mathbf{P}}_{D,m}^\intercal\right)^{-1}$, $\widetilde{\mathbf{c}}^\intercal$ is the first row of $\widetilde{\mathbf{Q}}_k$ and $\mathbf{b}$ is the first column of $\widetilde{\mathbf{P}}_m$.

In sequel we assume that $\mathbf{A}$, $\mathbf{D} \in \mathbb{R}^{d\times d}$, $\mathbf{b}$, $\mathbf{c}$ are in the balanced canonical form.