Subspace Identification From Classical Realization Methods

Daniel N. Miller and Raymond A. de Callafon

Dept. of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive 0411, La Jolla, CA 92093

Abstract: Subspace identification techniques are reinterpreted via classical realization theory to formulate a wide class of subspace identification methods. Re-formulating subspace identification in terms of a low rank decomposition of a weighted Hankel matrix allows special cases such as impulse-based and step-based input signals, but also realization based on arbitrary input signals and correlation functions. Ideas are illustrated with a simulation example.

1. INTRODUCTION

Of the various approaches to system identification, "subspace state-space system identification" methods – collectively referred to as 4SID methods – represent a powerful class of algorithms capable of providing consistent estimates of state-space models. The goal of subspace methods is to construct a realization of a system from the identification of the subspace spanned by the extended observability matrix. A reduced-order approximation of this matrix is subsequently determined via the singular value decomposition (SVD), and the mapping from input-to-state is solved for in a least-squares sense. Popular subspace methods include N4SID [Van Overschee and De Moor (1994)], CVA [Larimore (1990)], and MOESP [Verhaegen and Deprettere (1991)], which are presented in a unified framework in Van Overschee and De Moor (1996) and Katayama (2005).

Related to subspace identification methods is the realization algorithm of Kung (1978), which uses noise-corrupted estimates of the system Markov parameters to construct the system Hankel matrix. The SVD is then used to form a reduced-order approximation of the Hankel matrix, and the shift-invariant property of this matrix is used to solve for the system dynamics. A similar generalization of Kung's algorithm to arbitrary input-output data was previously presented in de Callafon et al. (2008).

In this paper, we generalize Kung's realization algorithm to arbitrary input-output data, formulating a subspace method with a direct connection to classical realization theory that may be used with input-output correlation functions as well. First, an overview of state-space realization from Markov parameters is provided, followed by a revised derivation of subspace methods. We then show that the oblique projection of subspace identification can be applied to any Markov-parameter convolution, resulting in a generalization of classical realization methods to arbitrary signals.

2. SYSTEM IDENTIFICATION VIA REALIZATION

2.1 Notation and Problem Formulation

Consider a discrete, linear, time-invariant, system described in state-space form as

$$x(t+1) = Ax(t) + Bu(t), \; x(0) = x_0$$
$$y(t) = Cx(t) + Du(t) + v(t) \quad (1)$$

with state vector $x \in \mathbb{R}^n$, input vector $u \in \mathbb{R}^m$, output vector $y \in \mathbb{R}^p$, and system matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$. The vector $v(t) \in \mathbb{R}^p$ is a zero-mean noise signal with constant variance. We restrict ourselves to systems that are both controllable and observable.

The input-output map can alternatively be described by the convolution operation

$$y(t) = \sum_{k=0}^{\infty} g(k)u(t-k) + v(t) \quad (2)$$

involving the system Markov parameters $g(k)$ defined as

$$g(k) = \begin{cases} D & k = 0 \\ CA^{k-1}B & k > 0 \end{cases}. \quad (3)$$

When applied to sequences of data, (2) may be expressed with two matrices mapping past and future input sequences, u_p and u_f respectively, onto a future output sequence y_f, as

$$y_f = Hu_p + Tu_f + v \quad (4)$$

in which

$$u_p = [u(0) \; u(-1) \; u(-2) \; \cdots]^T,$$
$$u_f = [u(1) \; u(2) \; u(3) \; \cdots]^T,$$
$$y_f = [y(1) \; y(2) \; y(3) \; \cdots]^T,$$
$$v = [v(1) \; v(2) \; v(3) \; \cdots]^T,$$

and

$$H = \begin{bmatrix} g(1) & g(2) & \cdots \\ g(2) & g(3) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \; T = \begin{bmatrix} g(0) & & 0 \\ g(1) & g(0) & \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

T is a block-Toeplitz matrix, and H is the familiar block-Hankel matrix of system Markov parameters [Van Der Veen et al. (1993)].

2.2 Deterministic Realization from Markov Parameters

The objective of the deterministic realization problem is to find the system order n and a state-space model (1) with respect to an arbitrary state basis given a finite number of system Markov parameters (3). A celebrated algorithm

given by Ho and Kalman (1966) uses the following two facts to obtain the system matrices (A, B, C) from H:

(1) H is the product of the extended observability matrix Γ and the extended controllability matrix Ω:

$$H = \Gamma\Omega = \begin{bmatrix} C & CA & CA^2 & \cdots \end{bmatrix}^T \begin{bmatrix} B & AB & A^2B & \cdots \end{bmatrix}.$$

(2) H exhibits a *shift-invariant* property such that should H be shifted up by p rows or to the left by m rows, the result is a block-Hankel matrix $\overrightarrow{H}$ with the property

$$\overrightarrow{H} = \Gamma A \Omega.$$

A controllable and observable system implies that Γ and Ω respectively have full column and full row rank n, making H a rank-n matrix. Thus, given a finite-dimensional pair $H \in \mathbb{R}^{N_1 \times N_2}$ and $\overrightarrow{H} \in \mathbb{R}^{N_1 \times N_2}$

$$H = \begin{bmatrix} g(1) & g(2) & \cdots & g(n_2) \\ g(2) & g(3) & \cdots & g(n_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ g(n_1) & g(n_1+1) & \cdots & g(n_1+n_2-1) \end{bmatrix} \quad (5)$$

$$\overrightarrow{H} = \begin{bmatrix} g(2) & g(3) & \cdots & g(n_2+1) \\ g(3) & g(4) & \cdots & g(n_2+2) \\ \vdots & \vdots & \ddots & \vdots \\ g(n_1+1) & g(n_1+2) & \cdots & g(n_1+n_2) \end{bmatrix}$$

in which $N_1 = n_1 \cdot p \geq n$ and $N_2 = n_2 \cdot m \geq n$, for any factorization of H

$$H = \Gamma\Omega$$
$$= \begin{bmatrix} C & CA & \cdots & CA^{n_1-1} \end{bmatrix}^T \begin{bmatrix} B & AB & \cdots & A^{n_2-1}B \end{bmatrix}$$

there exist a left inverse $\Gamma^\dagger$ of Γ and a right inverse $\Omega^\dagger$ of Ω such that A may be solved for as

$$A = \Gamma^\dagger \overrightarrow{H} \Omega^\dagger \quad (6)$$

in which $(\cdot)^\dagger$ denotes the Moore-Penrose psuedoinverse. C is subsequently taken as the first p rows of Γ and B from the first m columns of Ω. With D taken from $g(0)$, a complete, irreducible, state-space realization of (1) is obtained.

2.3 Realization From Noise-Corrupted Markov Parameters

If imperfect estimates of Markov parameters are used to construct an estimate $\hat{H}$ of (5), non-deterministic effects will likely cause $\hat{H}$ to have full rank. As shown in Kung (1978), both the system order n and a rank-n estimate of H can be computed from the SVD of $\hat{H}$:

$$\hat{H} = [U_n \ U_s] \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_s \end{bmatrix} \begin{bmatrix} V_n^T \\ V_s^T \end{bmatrix}$$

in which Σ_n and Σ_s are diagonal matrices containing the singular values ordered from largest to smallest, such that the first n singular values of $\hat{H}$ are contained in Σ_n and the remaining in Σ_s.

In the deterministic case, $\hat{H} = H$, $\text{rank}(\hat{H}) = n$, and $\Sigma_s = 0$. In the non-deterministic case, a "large" $\hat{H}$ will exhibit a relatively large difference in the value between the nth and the $(n+1)^{\text{th}}$ singular value. An estimate of n may be determined by choosing a threshold at which

the singular values decrease significantly, and a rank-n approximation of $\hat{H}$ is then given by

$$\hat{H}_n = U_n \Sigma_n V_n^T. \quad (7)$$

The decomposition of $\hat{H}_n$ in (7) also provides a convenient means of estimating the controllability and observability matrices as

$$\Gamma = U_n \Sigma_n^{1/2}, \qquad \Omega = \Sigma_n^{1/2} V_n^T,$$

with the expressions for the left inverse $\Gamma^\dagger$ and right inverse $\Omega^\dagger$ in (6) simplifying to

$$\Gamma^\dagger = \Sigma_n^{-1/2} U_n^T, \qquad \Omega^\dagger = V_n \Sigma_n^{-1/2}.$$

Many methods exist to generate state-space realizations with the above algorithm, notably Juang and Pappa (1985) and King et al. (1988). A disadvantage of these methods is that either a broad-band excitation signal or free-response data is required to provide reliable Markov parameter estimates.

3. SUBSPACE IDENTIFICATION

3.1 Elementary Data Equation

Subspace-based identification techniques formulate a lower-rank decomposition of the map from past-input to future-output without estimating the Markov parameters directly. The following differs slightly from common derivations in terms of matrix definitions to emphasize the equivalence with classical realization algorithms.

We return to the system description given in (4). Because (1) is an LTI system, (4) may be expanded column-wise to include arbitrarily many sequences of data so that the data-sequence vectors y_f, u_p, and u_f become data matrices Y_f, U_p, and U_f, respectively. In the realistic case in which we have access to only a single, finite sequence of input-output data, we may time-shift the data sequences to construct these data matrices. We also likely have no knowledge of the input sequence prior to our measured output, so we must separate the effects of the initial state x_0 on the future output into an auxiliary term $X_0 = \begin{bmatrix} Ax_0 & A^2x_0 & \cdots & A^{n_2}x_0 \end{bmatrix}$. The resulting equation for a sequence of $N = n_1 + n_2$ data points is

$$Y_f = HU_p + TU_f + \Gamma X_0 + V \quad (8)$$

in which

$$Y_f = \begin{bmatrix} y(1) & y(2) & \cdots & y(n_2) \\ y(2) & y(3) & \cdots & y(n_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(n_1) & y(n_1+1) & \cdots & y(n_1+n_2-1) \end{bmatrix}, \quad (9)$$

$$U_p = \begin{bmatrix} u(0) & u(1) & \cdots & u(n_2-1) \\ 0 & u(0) & \cdots & u(n_2-2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u(0) \end{bmatrix},$$

$$U_f = \begin{bmatrix} u(1) & u(2) & \cdots & u(n_2) \\ u(2) & u(3) & \cdots & u(n_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ u(n_1+1) & u(n_1+2) & \cdots & u(n_1+n_2) \end{bmatrix},$$

$$V = \begin{bmatrix} v(1) & v(2) & \cdots & v(n_2) \\ v(2) & v(3) & \cdots & v(n_2+1) \\ \vdots & \vdots & \ddots & \vdots \\ v(n_1) & v(n_1+1) & \cdots & v(n_1+n_2-1) \end{bmatrix},$$

$$T = \begin{bmatrix} g(0) & 0 & \cdots & 0 & 0 \\ g(1) & g(0) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ g(n_1-1) & g(n_1-2) & \cdots & g(0) & 0 \end{bmatrix} \quad (10)$$

With $N_1 = n_1 \cdot m$ and $N_2 = n_2 \cdot p$, the matrix U_p is a $N_2 \times N_2$ upper-triangular block-Toeplitz matrix, and T is a $N_1 \times N_1 + 1$ block-Toeplitz matrix. The zero column in T is included to provide a place-holder for working with a shifted version $\overrightarrow{T}$ of the same size without losing $g(0)$ in the proceeding section. The matrices U_f and V are block-Hankel matrices of size $N_1 + 1 \times N_2$ and $N_1 \times N_2$ respectively.

Note that the first data point in the matrices Y_f, U_f, and V is at $t = 1$, whereas in the traditional formulation the data begins at $t = 0$. A similar formulation of subspace identification with Hankel matrices was pursued in Vajk and Hetthessy (2005), but was not applied to the realization problem.

3.2 Subspace Identification via the Extended Observability Matrix

Subspace-based identification methods reconstruct the extended observability matrix Γ from input-output data, with the various methods differing primarily in how the system is then extracted from Γ. As such, the block-Hankel matrix H in (8) is factored into the extended observability matrix Γ and the extended controllability matrix Ω:

$$Y_f = \Gamma(\Omega U_p + X_0) + TU_f + V. \quad (11)$$

Note that the subspace equation (11) is traditionally written with a matrix of states X in place of $\Omega U_p + X_0$.

Remark 1. With $x_0 = 0$ and $u(t)$ a unit-pulse input applied at $t = 0$, (11) simplifies to $Y_f = H + V$, and the subspace identification problem reduces to a realization from Markov parameter estimates [Van Overschee and De Moor (1996)].

To isolate the extended observability matrix Γ in (11), the effect of U_f is removed by projecting Y_f onto the null space of U_f, such that

$$Y_f \Pi_{U_f^\perp} = \Gamma(\Omega U_p + X_0)\Pi_{U_f^\perp} + V\Pi_{U_f^\perp}$$

in which

$$\Pi_{U_f^\perp} = I - U_f^T (U_f U_f^T)^\dagger U_f \quad (12)$$

The projection of Y_f is subsequently weighted by matrices W_1 and W_2 such that

$$W_1 Y_f \Pi_{U_f^\perp} W_2 = W_1 \Gamma(\Omega U_p + X_0)\Pi_{U_f^\perp} W_2 \\ + W_1 V \Pi_{U_f^\perp} W_2$$

in which W_1 and W_2 are chosen to be rank-preserving and such that $W_1 V \Pi_{U_f^\perp} W_2 \to 0$ as the number of samples $N \to \infty$. Explorations of the role of W_1 and W_2 as well as suggestions for their contents can be found in Van Overschee and De Moor (1996) and Favoreel et al. (2000). Subsequently, the SVD can be applied to $W_1 Y_f \Pi_{U_f^\perp} W_2$

and, as in Kung's algorithm, the system order n is chosen by examining the singular values. The decomposition is used to compute a rank-n estimate of Γ, the most common factorization of which is

$$W_1\Gamma = U_n \Sigma_n^{1/2}$$
$$(\Omega U_p + X_0)\Pi_{U_f^\perp} W_2 = \Sigma_n^{1/2} V_n^T$$

The projection operation makes it impossible to guarantee that Ω can be successfully determined from the above factorization. An LQ-decomposition-based stochastic realization algorithm presented in Tanaka and Katayama (2006) does extract Ω from a similar decomposition and factorization but does not include a projection step due to the nature of the problem considered. Instead, subspace algorithms take one of the following approaches: either treat $\Sigma_n^{1/2} V_n^T$ as a space of system states, or identify the A and C matrices from the shift-invariant property of Γ. Whichever method is chosen, the remaining unknown parameters are linear in (1) and can be computed via a least squares solution.

4. SUBSPACE IDENTIFICATION VIA SHIFT-INVARIANCE OF OUTPUT DATA

4.1 Shifted Elementary Data Equation

The shift-invariance of the Hankel matrices in (5) can be extended to the projected output term $Y_f \Pi_{U_f^\perp}$, allowing for the union of subspace-based and realization-based approaches into a single algorithm. Additionally, by generalizing the approach of subspace identification to Hankel operators, the applications of subspace techniques may be extended to any arbitrary situation in which the convolution of Markov parameters may be found.

We begin by reexamining the subspace equation (11) and define $\overrightarrow{Y}_f$ as Y_f shifted by a single column to the left, thus now including $y(N)$ $(N = n_1 + n_2)$:

$$\overrightarrow{Y}_f = \begin{bmatrix} y(2) & y(3) & \cdots & y(n_2+1) \\ y(3) & y(4) & \cdots & y(n_2+2) \\ \vdots & \vdots & \vdots & \vdots \\ y(n_1+1) & y(n_1+2) & \cdots & y(n_1+n_2) \end{bmatrix}.$$

Performing this column-wise shift allows one to write a shifted data equation

$$\overrightarrow{Y}_f = \overrightarrow{H}U_p + \overrightarrow{T}U_f + \overrightarrow{\Gamma}X_0 + \overrightarrow{V} \quad (13)$$

where $\overrightarrow{H}$ is given in (5) and

$$\overrightarrow{T} = \begin{bmatrix} g(1) & g(0) & 0 & \cdots & 0 \\ g(2) & g(1) & g(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g(N_1) & g(N_1-1) & \cdots & g(1) & g(0) \end{bmatrix}, \ \overrightarrow{\Gamma} = \begin{bmatrix} CA \\ \vdots \\ CA^{n_1} \end{bmatrix}.$$

The shifted data equation can also be written as

$$\overrightarrow{Y}_f = \overrightarrow{\Gamma}(\Omega U_p + X_0)\overrightarrow{T}U_f + \overrightarrow{V} \quad (14)$$

by factoring the shifted Hankel matrix $\overrightarrow{H}$ into the shifted extended observability matrix $\overrightarrow{\Gamma}$ and the extended controllability matrix Ω. The data equation given in (8), (11) and the shifted data equation in (13), (14) form the basis of the following subspace-based identification algorithms.

4.2 Realization from Step-Response Data

Instead of applying a unit-pulse as in Remark 1, consider the more practical situation of applying a unit-step input with initial state $x_0 = 0$. The matrix U_p is then an upper-triangular unity matrix, and the matrix product TU_f is given by

$$TU_f = \begin{bmatrix} g(0) & g(0) & \cdots & g(0) \\ g(0)+g(1) & g(0)+g(1) & \cdots & g(0)+g(1) \\ \vdots & \vdots & & \vdots \\ \sum_{i=0}^{L} g(i) & \sum_{i=0}^{L} g(i) & \cdots & \sum_{i=0}^{L} g(i) \end{bmatrix}$$

which is equivalent to the noise-free step response data

$$TU_f = \begin{bmatrix} y(0) & y(0) & \cdots & y(0) \\ y(1) & y(1) & \cdots & y(1) \\ \vdots & \vdots & & \vdots \\ y(L) & y(L) & \cdots & y(L) \end{bmatrix}. \tag{15}$$

In the deterministic case, the *weighted* Hankel matrix $R = HU_p$ can be computed from (8)

$$R = HU_p = Y_f - TU_f$$

while a similar shifted version $\overrightarrow{R}$ can be computed from (13) [de Callafon (2003)]. The following results hold for R and $\overrightarrow{R}$ in the deterministic case:

- If $u(0) \neq 0$, the square $N_2 \times N_2$ matrix U_p has full rank. Consequently, $\mathrm{rank}(R) = \mathrm{rank}(H) = n$, and R has a decomposition $R = R_1 R_2$ in which R_1 and R_2 have respectively full column rank n and full row rank n. In the case of noise-corrupted observations, a rank-n approximation of R can be found via its SVD.
- Since $R = HU_p$ is a post-multiplication of H, R exhibits the same shift-invariant property as H, and therefore $\overrightarrow{R} = R_1 A R_2$ can be used to recover A via left and right inverses taken from the decomposition $R = R_1 R_2$.

The above observations imply that a realization algorithm similar to Kung's can be formulated using R instead of H.

Remark 2. Given the output $y(t)$ of a unit-step response, applying the realization algorithm to $R = Y_f - TU_f$ will yield significantly better results than applying the algorithm to an estimate of H obtained by *differentiating* the step response measurements, as doing so would amplify high-frequency noise and increase the variance of the estimated parameters.

4.3 Realization from Arbitrary Input-Output Data

Consider (11) and its shifted version (14) for an arbitrary $u(t)$ and an unknown x_0. Projecting both Y_f and $\overrightarrow{Y}_f$ onto the null-space of U_f results in

$$\begin{aligned} Y_f \Pi_{U_f^\perp} &= \Gamma(\Omega U_p + X_0)\Pi_{U_f^\perp} + V\Pi_{U_f^\perp} \\ \overrightarrow{Y}_f \Pi_{U_f^\perp} &= \overrightarrow{\Gamma}(\Omega U_p + X_0)\Pi_{U_f^\perp} + \overrightarrow{V}\Pi_{U_f^\perp} \end{aligned} \tag{16}$$

If $x_0 = 0$, $\Gamma\Omega$ and $\overrightarrow{\Gamma}\Omega$ may be collected as H and $\overrightarrow{H}$, respectively.

The goal is to apply Kung's algorithm to the pair $Y_f\Pi_{U_f^\perp}$ and $\overrightarrow{Y}_f\Pi_{U_f^\perp}$. Doing so requires that the rank of H be

preserved in equation above. With this in mind, the following observations regarding the rank of each term can be made:

- If $\mathrm{rank}(U_f) = k \leq \min(N_1+1, N_2)$, then $\mathrm{rank}(\Pi_{U_f^\perp}) = N_2 - k$ due to the fact that $\Pi_{U_f^\perp}$ is an $N_2 \times N_2$ matrix and is defined as a projection onto the null-space of U_f.
- Without loss of generality, we can assume $u(0) \neq 0$ and $\mathrm{rank}(U_p) = N_2$ (U_p is full rank).
- Because $\mathrm{rank}(H) = n \leq \min(N_1, N_2)$, the rank conditions on U_p imply that $\mathrm{rank}(HU_p\Pi_{U_f^\perp}) \leq \min(n, \mathrm{rank}(\Pi_{U_f^\perp}))$.
- If $V = 0$ (the deterministic case), $\mathrm{rank}(HU_p\Pi_{U_f^\perp}) = \mathrm{rank}(Y_f\Pi_{U_f^\perp})$. If $V \neq 0$, then $\mathrm{rank}(HU_p\Pi_{U_f^\perp}) \leq \mathrm{rank}(Y_f\Pi_{U_f^\perp})$.

Thus the constraint $\mathrm{rank}(\Pi_{U_f^\perp}) \geq n$ is sufficient to preserve the rank of H. If $N_2 > N_1 + n$, $\mathrm{rank}(U_f) = k \leq N_1$, and the following result is obtained:

$$\mathrm{rank}(\Pi_{U_f^\perp}) = N_2 - k > N_1 + n - k \geq N_1 + n - N_1 = n$$

From this, we can conclude that the use of "wide" matrices Y_f and U_p in (16) for which $N_2 > N_1 + n$ will preserve the rank of H.

As with other subspace methods, a proper choice of weighting matrices W_1 and W_2 results in the eventual elimination of the noise terms such that $W_1 V\Pi_{U_f^\perp} W_2 \to 0$ and $W_1 \overrightarrow{V}\Pi_{U_f^\perp} W_2 \to 0$ as $N \to \infty$, reducing (16) to

$$\begin{aligned} W_1 Y_f \Pi_{U_f^\perp} W_2 &= W_1 H U_p \Pi_{U_f^\perp} W_2 \\ W_1 \overrightarrow{Y}_f \Pi_{U_f^\perp} W_2 &= W_1 \overrightarrow{H} U_p \Pi_{U_f^\perp} W_2 \end{aligned} \tag{17}$$

Assume that $u(t)$ is persistently exciting and that the matrices W_1 and W_2 are rank-preserving, and let Q_1 and Q_2 be any factorization $Q_1 Q_2 = W_1 Y_f \Pi_{U_f^\perp} W_2$ in which Q_1 has n columns with full column rank and in which Q_2 has n rows with full row rank. There then exist a Γ and Ω such that

$$Q_1 = W_1\Gamma, \qquad Q_2 = \Omega U_p \Pi_{U_f^\perp} W_2, \qquad H = \Gamma\Omega.$$

Non-deterministic effects, however, require that Q_1 and Q_2 instead be factored from a rank-n approximation of $W_1 Y_f \Pi_{U_f^\perp} W_2$. As in the previous approaches, the SVD is a useful tool for computing estimates of Q_1 and Q_2.

From the decomposition

$$W_1 Y_f \Pi_{U_f^\perp} W_2 = [U_n \ U_s] \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_s \end{bmatrix} \begin{bmatrix} V_n^T \\ V_s^T \end{bmatrix}$$

the system order n is determined by examining the singular values for some threshold between large and small values. Having decided on an appropriate system order n, Q_1 and Q_2 are taken to be

$$Q_1 = U_n \Sigma_n^{1/2} \qquad Q_2 = \Sigma_n^{1/2} V_n^T$$

Substitution of (6) into (17) then results in

$$W_1 \overrightarrow{H} U_p \Pi_{U_f^\perp} W_2 = W_1 \Gamma A \Omega U_p \Pi_{U_f^\perp} W_2 = Q_1 A Q_2$$

Because Q_1 has full column rank, it has a left-inverse $Q_1^\dagger$, and because Q_2 has full row rank, it has a right-inverse $Q_2^\dagger$, so that A may be solved for via

$$A = Q_1^\dagger W_1 \overrightarrow{Y}_f \Pi_{U_f^\perp} W_2 Q_2^\dagger$$

The matrix C can then be estimated as the first p rows of Q_1. With A and C known, B and D can then be solved for via least-squares methods.

Remark 3. With $x_0 = 0$ and $u(t)$ a unit-pulse input applied at $t = 0$, $U_f = 0_{N_1 \times N_2}$, $\Pi_{U_f^\perp} = I_{N_2 \times N_2}$, and (13) simplifies to $Y_f = H + V$. With $W_1 = W_2 = I$, the above algorithm again reduces to a realization from Markov parameter estimates.

The reinterpretation of the matrix product $W_1 Y_f \Pi_{U_f^\perp} W_2$ in terms of Hankel and Toeplitz operators allows us to apply subspace techniques in a more general framework. It is often the case that with large datasets, computing the singular value decomposition of $W_1 Y_f \Pi_{U_f^\perp} W_2$ can become computationally infeasible. Alternatively, we may seek out other relationships in which the Markov parameters appear and apply the same techniques. As a final note it should be mentioned that the above algorithm does not guarantee a stable state space model or positive real results for stochastic models, as it is subject to the same limitations of traditional subspace identification algorithms examined in Lindquist and Picci (1996).

4.4 Realization via Correlation Functions

Another situation in which the Markov-parameter convolution can be found is the relationship between the autocorrelation function of the input signal $R_u(\tau)$ and the cross-correlation of the output with the input $R_{yu}(\tau)$ which can be shown [Ljung (1999)] to satisfy the relation

$$R_{yu}(\tau) = \sum_{i=0}^{\infty} g(i) R_u(i - \tau).$$

By using correlation function estimates

$$R_u^N(\tau) = \frac{1}{N} \sum_{t=0}^{N} u(t) u(t - \tau), \quad R_{yu}^N(\tau) = \frac{1}{N} \sum_{t=0}^{N} y(t) u(t - \tau)$$

the effect noise for large data sets can be reduced while minimizing the size of the matrices on which the SVD is applied. This is due to the fact that

$$\lim_{N \to \infty} R_{vu}^N(\tau) = 0 \qquad (18)$$

provided the noise $v(t)$ is uncorrelated with the input $u(t)$.

Similar to the way in which the convolution sum of (2) was written in matrix form and separated into past and future components, the cross-correlation of $y(t)$ and $u(t)$ can be separated into operations on past and future sequences of correlation functions, but with respect to τ instead of t. The result is

$$R_{yu,f} = H R_{u,p} + T R_{u,f} + R_{vu} \qquad (19)$$

in which H and T are given in (5) and (10) respectively and $R_{yu,f}$ is given by

$$\begin{bmatrix} R_{yu}^N(1) & R_{yu}^N(2) & \cdots & R_{yu}^N(N_2) \\ R_{yu}^N(2) & R_{yu}^N(3) & \cdots & R_{yu}^N(N_2 + 1) \\ \vdots & \vdots & \vdots & \vdots \\ R_{yu}^N(N_1) & R_{yu}^N(N_1 + 1) & \cdots & R_{yu}^N(N_1 + N_2 - 1) \end{bmatrix} \qquad (20)$$

and

$$R_{u,p} = \begin{bmatrix} R_u^N(0) & R_u^N(1) & \cdots & R_u^N(N_2 - 1) \\ 0 & R_u^N(0) & \cdots & R_u^N(N_2 - 2) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & R_u^N(0) \end{bmatrix},$$

$$R_{u,f} = \begin{bmatrix} R_u^N(1) & R_u^N(2) & \cdots & R_u^N(N_2) \\ R_u^N(2) & R_u^N(3) & \cdots & R_u^N(N_2 + 1) \\ \vdots & \vdots & \vdots & \vdots \\ R_u^N(N_1 + 1) & R_u^N(N_1 + 2) & \cdots & R_u^N(N_1 + N_2) \end{bmatrix}.$$

R_{vu} is the matrix generated from shifting sequences of $R_{vu}(\tau)$ with τ. A projection similar to the one in (12) can then be applied such that

$$R_{yu,f} \Pi_{R_{u,f}^\perp} = H R_{u,p} \Pi_{R_{u,f}^\perp} + R_{vu} \Pi_{R_{u,f}^\perp} \qquad (21)$$

in which

$$\Pi_{R_{u,f}^\perp} = I - R_{u,f}^T (R_{u,f} R_{u,f}^T)^\dagger R_{u,f} \qquad (22)$$

A minimal-rank estimation of the left hand side of (21) may then be found via its SVD, but with the matrices much smaller in size than what would otherwise have been necessary. The resulting decomposition

$$R_{yu,f} \Pi_{R_{u,f}^\perp} = [U_n \; U_s] \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_s \end{bmatrix} \begin{bmatrix} V_n^T \\ V_s^T \end{bmatrix} \qquad (23)$$

is then used to estimate the system matrix A as

$$A = \Sigma_n^{-1/2} U_n^T \overrightarrow{R}_{yu,f} \Pi_{R_{u,f}^\perp} V_n \Sigma_n^{-1/2} \qquad (24)$$

in which $\overrightarrow{R}_{yu}$ is the matrix R_{yu} shifted by one unit of τ. The matrix C may then be determined from the top p rows of $U_n \Sigma_n^{1/2}$, and B and D may be solved for in a least-squares sense using any arbitrary (small) sequence of data.

Remark 4. If the input signal $\{u(t)\}$ is white noise, $R_{u,p} \to I$ and $R_{u,f} \to 0$ as $N \to \infty$. Therefore $R_{yu,f} \to H$, and the algorithm once more reduces into realization from Markov parameter estimates.

Remark 5. The averaging effects caused by first estimating the correlation functions may significantly reduce the size of the matrices required in the projection and decomposition steps of the subspace identification algorithm, allowing for more efficient processing of large data sets.

4.5 Simulation Example

To demonstrate the proposed subspace-based identification methods that use the shift-invariance property for both input-output data and the auto- and cross-correlation functions, we generate a random 10th order stable single-input, single-output discrete-time system $G(q)$ in Matlab and produce a pseudo-random binary input $u(t)$ and a noise disturbed output measurement $y_{nf}(t) = G(q)u(t)$, $y(t) = y_{nf}(t) + v(t)$. We produce $N = 2048$ data points and the noise $v(t)$ is a zero-mean white noise with a variance normalized to 1/10th of the variance of the noise free output $y_{nf}(t)$.

We create matrices Y_f and U_f in (8) with the size $N_1 = 20$, $N_2 = 1024$ to use all data points, and matrices $R_{yu,f}$ and $R_{u,f}$ in (19) of only $N_1 = 20$, $N_2 = 40$ to be able to estimate at least a 20th order model. Choosing weighting matrices $W_1 = W_2 = I$, the singular value plots for $Y_f \Pi_{U_f^\perp}$, $R_{yu,f} \Pi_{R_{u,f}^\perp}$, and the original Hankel matrix H,

are shown in Figure 1 after normalization so that the plots can be compared. The additional averaging inherent in the auto- and cross-correlations enables a better separation between the large and small singular values in $R_{yu,f}\Pi_{R_{u,f}^{\perp}}$, whereas the singular value plot of $Y_f\Pi_{U_f^{\perp}}$ levels due to the noise. The singular value plot shows a 7th order model can be estimated and the result has been depicted in Figure 2.
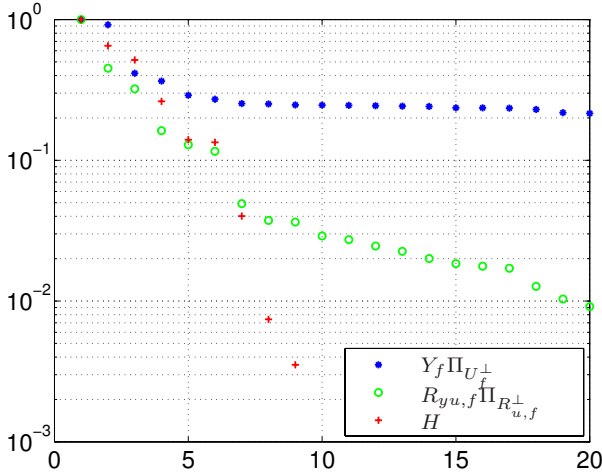

Fig. 1. Singular value plot (normalized) of $Y_f\Pi_{U_f^{\perp}}$ given in (9), (12) and $R_{yu,f}\Pi_{R_{u,f}^{\perp}}$ given in (20), (22) for the same noisy data set. For comparison, the singular values of H in (5) for noise free data has been plotted.
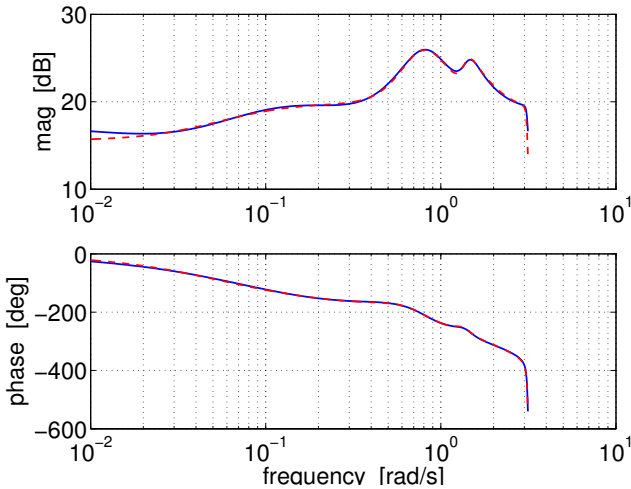

Fig. 2. Bode plot of original 10th order system $G(q)$ (solid) and identified 7th order state space model (dashed).

5. CONCLUSION

Subspace identification techniques are interpreted in the framework of classical realization theory by applying the shift invariant property of the Hankel matrix directly to input-output data or auto-/cross-correlation functions. It is shown that matrix computations collapse to classical realization techniques for special experimental conditions. In addition, realization algorithms are formulated based on step response data or correlation functions.

REFERENCES

R. A. de Callafon. Estimating parameters in a lumped parameter system with first principle modeling and dynamic experiments. In *System Identification, 2003. 13th IFAC Symposium on*, pages 1613–1618, 2003.

R. A. de Callafon, B. Moaveni, J. P. Conte, X. He, and E. Udd. General realization algorithm for modal identification of linear dynamic systems. *Journal of Engineering Mechanics*, 134(9):712–722, 2008.

W. Favoreel, B. De Moor, and P. Van Overschee. Subspace state space system identification for industrial processes. *Journal of Process Control*, 10(2-3):149–155, April 2000.

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

J. N. Juang and R. S. Pappa. An eigensystem realization algorithm (ERA) for modal parameter identification and model reduction. *JPL Proc. of the Workshop on Identification and Control of Flexible Space Structures*, 3:299–318, April 1985.

T. Katayama. *Subspace Methods for System Identification.* Communications and Control Engineering. Springer-Verlag, 2005.

A. M. King, U. B. Desai, and R. E. Skelton. A generalized approach to q-Markov covariance equivalent realizations for discrete systems. *Automatica*, 24(4):507–515, July 1988.

S. Y. Kung. A new identification and model redaction algorithm via singular value decomposition. In *12th Asilomar Conference on Circuits, Systems and Computers*, pages 705–714, 1978.

W. E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *Decision and Control, 1990., Proceedings of the 29th IEEE Conference on*, pages 596–604 vol.2, 1990.

A. Lindquist and G. Picci. Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica*, 32(5):709–733, May 1996.

L. Ljung. *System Identification: Theory for the User*. PTR Prentice Hall Information and System Sciences. Prentice Hall PTR, Upper Saddle River, NJ, 2nd edition, 1999.

H. Tanaka and T. Katayama. A stochastic realization algorithm via block LQ decomposition in hilbert space. *Automatica*, 42(5):741–746, May 2006.

I. Vajk and J. Hetthessy. Subspace identification methods: review and re-interpretation. In *Control and Automation, 2005. ICCA '05. International Conference on*, volume 1, pages 113–118 Vol. 1, 2005.

A. Van Der Veen, E. F. Deprettere, and A. L. Swindlehurst. Subspace-based signal analysis using singular value decomposition. *Proceedings of the IEEE*, 81(9):1277–1308, 1993.

P. Van Overschee and B. De Moor. N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.

P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementatin, Applications.* Kluwer Academic Publishers, 1996.

M. Verhaegen and E. Deprettere. A fast, recursive MIMO state space model identification algorithm. In *Decision and Control, 1991., Proceedings of the 30th IEEE Conference on*, pages 1349–1354 vol.2, 1991.