Proceedings of the 18th World Congress
The International Federation of Automatic Control
Milano (Italy) August 28 - September 2, 2011

Efficient Identification of Input Dynamics for Correlation Function-Based Subspace Identification

Daniel N. Miller *, **Raymond A. de Callafon** *

* *University of California San Diego*
(email: d6miller@ucsd.edu, callafon@ucsd.edu)

Abstract: The methods of subspace system identification are extended to correlation function estimates, explicitly addressing the increase in computational difficulty of identifying input-to-state dynamics when correlation function estimates are used in place of input-output data for multivariable identification problems. It is shown that the regressor used to solve a common least-squares problem when identifying input-to-state dynamics is the state sequence of a dual system of dynamics that have already been estimated. A new method of computing the regressor is presented that dramatically improves the computational efficiency of estimating the input-to-state dynamics when signals of high dimension are used. A simulation example demonstrates the effectiveness of the method for both input-output data and correlation function estimates.

1. INTRODUCTION

Subspace identification methods form a popular class of algorithms for identifying linear, state-space models from experimental input-output data without requiring nonlinear optimization. Overviews of subspace identification can be found in Van Overschee and De Moor [1996a], Ljung [1999], Katayama [2005], and Verhaegen and Verdult [2007]. Although usually concerned with time-domain identification, subspace methods extend naturally to other domains. In Van Overschee and De Moor [1996b] and McKelvey et al. [1996], subspace methods were extended to frequency-response function estimates, and in Di Ruscio [1995], Miller and de Callafon [2009], and Miller and de Callafon [2010], subspace identification methods were extended to correlation function estimates.

A consequence of extending subspace methods to correlation function estimates is that the input-output signals may become matrix-valued when correlation function estimates are generated from multivariable data, effectively increasing the dimension of the signals used. Although the size of the matrices needed to estimate the range of the extended observability matrix is reduced, this complicates the least-squares identification of the input-to-state dynamics commonly used in subspace identification; the regressor needed to solve the linear-least-squares problem becomes large and difficult to compute using standard formulas currently available in the literature.

Subspace methods are particularly appealing in situations when the measured signals are of high dimensions, since alternative identification methods based on nonlinear optimizations frequently suffer from convexity issues, yet memory availability is often a limiting factor when using multivariable data sets with subspace methods. Difficulties with increased dimension have thus far significantly limited the applications to which the authors have successfully applied a correlation function-based approach.

To improve the efficiency of the linear-least-squares identification of input-to-state dynamics, we show that the regressor can be formed from a set of state-sequences of a dual system, which can be described by the state-to-output dynamics that have already been estimated. This alternative formulation is first demonstrated using standard input-output data and then extended to correlation function estimates. The dramatic reduction in size of the matrices used to solve for input-to-state dynamics greatly increases the maximum possible dimensions of the input and output signals as well as the maximum the size of the data sets.

2. PRELIMINARIES

Consider a linear, time-invariant, discrete-time system described by the state-space equations

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) + v(t) \end{aligned} \tag{1}$$

which relate the input $u(t) \in \mathbb{R}^{n_u}$ to the state $x(t) \in \mathbb{R}^n$ and the output $y(t) \in \mathbb{R}^{n_y}$ in terms of the constant matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n_u}$, $C \in \mathbb{R}^{n_y \times n}$, and $D \in \mathbb{R}^{n_y \times n_u}$. Added to the output is a possibly colored noise signal $v(t) \in \mathbb{R}^{n_y}$, assumed to be the realization of a stationary stochastic process that may or may not share dynamics with the system described by (A, B, C, D).

If $u(t)$ is selected to be quasi-stationary, then the quasi-stationary property of $v(t)$ will result in a quasi-stationary $y(t)$ [Ljung, 1999]. This guarantees that the auto-correlation function

$$R_u(\tau) = E\left[u(t+\tau)u(t)^T\right] \in \mathbb{R}^{n_u \times n_u}$$

and the cross-correlation functions

$$R_{yu}(\tau) = E\left[y(t+\tau)u(t)^T\right] \in \mathbb{R}^{n_y \times n_u}$$

and

$$R_{vu}(\tau) = E\left[v(t+\tau)u(t)^T\right] \in \mathbb{R}^{n_y \times n_u}$$

will exist. If we define the cross-correlation of the state with the input as

$$R_{xu}(\tau) = E\left[x(t+\tau)u(t)^T\right] \in \mathbb{R}^{n\times n_u},$$

then the correlation functions may be expressed in terms of the state-space matrices (A, B, C, D) as

$$\begin{aligned} R_{xu}(\tau+1) &= AR_{xu}(\tau) + BR_u(\tau) \\ R_{yu}(\tau) &= CR_{xu}(\tau) + DR_u(\tau) + R_{vu}(\tau). \end{aligned}$$

Additionally, if the input and noise signals are uncorrelated (the system is operating in open-loop), then

$$R_{vu}(\tau) = 0 \quad \forall \quad \tau.$$

A thorough discussion of quasi-stationary signals and the relationships of correlation functions may be found in Ljung [1999].

3. IDENTIFICATION OF STATE-TO-OUTPUT DYNAMICS

The following is a brief review of subspace identification methods that identify the parameters A and C from an estimated range of the extended observability matrix. We then extend the basic method to correlation function estimates and discuss some of the advantages of doing so.

3.1 Identification from Input-Output Data

Traditional subspace identification involves creating block-Hankel matrices of input and output data, each with i block rows and l block columns, forming

$$U = \begin{bmatrix} u(0) & u(1) & \cdots & u(l-1) \\ u(1) & u(2) & \cdots & u(l) \\ \vdots & \vdots & & \vdots \\ u(i-1) & u(i) & \cdots & u(i+l-2) \end{bmatrix} \in \mathbb{R}^{in_u\times l}$$

and

$$Y = \begin{bmatrix} y(0) & y(1) & \cdots & y(l-1) \\ y(1) & y(2) & \cdots & y(l) \\ \vdots & \vdots & & \vdots \\ y(i-1) & y(i) & \cdots & y(i+l-2) \end{bmatrix} \in \mathbb{R}^{in_y\times l}.$$

These data matrices are related to each other in terms of a block-Toeplitz matrix

$$T = \begin{bmatrix} D & & & & \\ CB & D & & & \\ CAB & CB & D & & \\ \vdots & \vdots & \vdots & \ddots & \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \cdots & D \end{bmatrix} \in \mathbb{R}^{in_y\times in_u},$$

an extended observability matrix

$$\Gamma = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{bmatrix} \in \mathbb{R}^{in_y\times n},$$

a matrix of states

$$X = \begin{bmatrix} x(0) & x(1) & \cdots & x(l-1) \end{bmatrix},$$

and a block-Hankel matrix of noise data

$$V = \begin{bmatrix} v(0) & v(1) & \cdots & v(l-1) \\ v(1) & v(2) & \cdots & v(l) \\ \vdots & \vdots & & \vdots \\ v(i-1) & v(i) & \cdots & v(i+l-2) \end{bmatrix} \in \mathbb{R}^{in_y\times l},$$

as

$$Y = \Gamma X + TU + V. \tag{2}$$

Multiplication of Y on the right by the projector matrix

$$\Pi = I - U^T\left(UU^T\right)^{-1}U$$

will project Y onto the null space of U, resulting in

$$Y\Pi = \Gamma X\Pi + V\Pi. \tag{3}$$

A rank-n approximation of Γ is calculated via the singular-value decomposition, and least-squares estimates $\hat{A}$ and $\hat{C}$ are then found from the structure of Γ. Specifically,

$$\hat{C} = \Gamma(1:n_y, :)$$

and

$$\hat{A} = \left(\Gamma(1:(i-1)n_y, :)\right)^{\dagger}\Gamma(n_y+1:in_y, :)$$

where MATLAB-style indexing notation has been used, and $(\cdot)^{\dagger}$ represents the Moore-Penrose pseudoinverse. Detailed overviews and implementation aspects of subspace identification may be found in many places, particularly those references cited in the introduction.

3.2 The Effects of Colored Noise on System Estimates

A potential issue with using input-output data to estimate the extended observability matrix in this manner arrises when the spectrum of the noise process $v(t)$ cannot be represented by additive white noise on the state equation (1). To demonstrate this, suppose $v(t)$ can be represented as a filtered white-noise signal $e(t)$, where

$$\begin{aligned} x_v(t+1) &= A_v x_v(t) + Ke(t) \\ v(t) &= C_v x_v(t) + e(t), \end{aligned} \tag{4}$$

the pair (A_v, C_v) is observable, and A_v contains at least one eigenvalue different from those of A. Letting

$$T_v = \begin{bmatrix} I & & & \\ C_vK & I & & \\ \vdots & \vdots & \ddots & \\ C_vA_v^{i-2}K & C_vA_v^{i-3}K & \cdots & I \end{bmatrix},$$

$$\Gamma_v = \begin{bmatrix} C_v \\ C_vA_v \\ \vdots \\ C_vA_v^{i-1} \end{bmatrix}, \quad X_v = \begin{bmatrix} x_v(0) & x_v(1) & \cdots & x_v(l-1) \end{bmatrix},$$

and

$$Z = \begin{bmatrix} e(0) & e(1) & \cdots & e(l-1) \\ e(1) & e(2) & \cdots & e(l) \\ \vdots & \vdots & & \vdots \\ e(i-1) & e(i) & \cdots & e(i+l-2) \end{bmatrix} \in \mathbb{R}^{in_y\times l},$$

the projected data-matrix equation (3) becomes

$$Y\Pi = \Gamma X\Pi + \left(\Gamma_v X_v + T_v Z\right)\Pi.$$

Thus it is usually impossible to distinguish the range of Γ from the range of Γ_v and T_v when finding a low-rank approximation of $\Gamma X\Pi$. The only way to consistently estimate the eigenvalues of A will be to artificially choose the system order high enough so that the eigenvalues of both A and A_v appear in $\hat{A}$. Although a subsequent estimation of B and D will likely place blocking zeros [Zhou et al., 1995] near the eigenvalues of A_v, there is no way to conclude if these zeros are masking the noise or are part of the system dynamics.

3.3 Identification from Correlation Function Estimates

Because the signals $u(t)$ and $y(t)$ share the same dynamics as $R_u(\tau)$ and $R_{yu}(\tau)$, we may replace the input-output data $u(t)$ and $y(t)$ in the identification algorithm with $R_u(\tau)$ and $R_{yu}(\tau)$. Correlation-function data matrices

$$R_U = \begin{bmatrix} R_u(\tau_0) & R_u(1+\tau_0) & \cdots & R_u(l-1+\tau_0) \\ R_u(1+\tau_0) & R_u(2+\tau_0) & \cdots & R_u(l+\tau_0) \\ \vdots & \vdots & & \vdots \\ R_u(i-1+\tau_0) & R_u(i+\tau_0) & \cdots & R_u(i+l-2+\tau_0) \end{bmatrix} \in \mathbb{R}^{in_u \times ln_u},$$

$$R_{YU} = \begin{bmatrix} R_{yu}(\tau_0) & R_{yu}(1+\tau_0) & \cdots & R_{yu}(l-1+\tau_0) \\ R_{yu}(1+\tau_0) & R_{yu}(2+\tau_0) & \cdots & R_{yu}(l+\tau_0) \\ \vdots & \vdots & & \vdots \\ R_{yu}(i-1+\tau_0) & R_{yu}(i+\tau_0) & \cdots & R_{yu}(i+l-2+\tau_0) \end{bmatrix} \in \mathbb{R}^{in_y \times ln_u},$$

$$R_{VU} = \begin{bmatrix} R_{vu}(\tau_0) & R_{vu}(1+\tau_0) & \cdots & R_{vu}(l-1+\tau_0) \\ R_{vu}(1+\tau_0) & R_{vu}(2+\tau_0) & \cdots & R_{vu}(l+\tau_0) \\ \vdots & \vdots & & \vdots \\ R_{vu}(i-1+\tau_0) & R_{vu}(i+\tau_0) & \cdots & R_{vu}(i+l-2+\tau_0) \end{bmatrix} \in \mathbb{R}^{in_y \times ln_u},$$

and

$$R_{XU} = [R_{xu}(\tau_0) \; R_{xu}(1+\tau_0) \; \cdots \; R_{xu}(l-1++\tau_0)] \in \mathbb{R}^{ds},$$

in which τ_0 is some starting value of τ possibly < 0, then replace the data matrices in the original identification procedure, transforming (2) into

$$R_{YU} = \Gamma R_{XU} + T R_U + R_{VU}.$$

If $R_{vu}(\tau) = 0$, then $R_{VU} = 0$, and this becomes

$$R_{YU} = \Gamma R_{XU} + T R_U.$$

Multiplication of R_{YU} on the right by the null-space-projection matrix

$$\Pi_R = I - R_U^T \left(R_U R_U^T\right)^{-1} R_U,$$

will remove the effects of the future auto-correlation functions R_U, resulting in

$$R_{YU}\Pi_R = \Gamma R_{XU}.$$

And thus the range of Γ is isolated, regardless of the contents of the noise.

Because the functions $R_u(\tau)$ and $R_{yu}(\tau)$ are unknown, we replace them with estimates taken from N samples of input-output data

$$\begin{aligned} \hat{R}_u(\tau) &= \frac{1}{N} \sum_{t=0}^{N-\tau-1} u(t+\tau)u(t)^T \\ \hat{R}_{yu}(\tau) &= \frac{1}{N} \sum_{t=0}^{N-\tau-1} y(t+\tau)u(t)^T, \end{aligned} \tag{5}$$

which, under the quasi-stationary assumptions of $u(t)$ and $y(t)$ and the assumption that $u(t)$ and $v(t)$ are uncorrelated, will converge to the true functions $R_u(\tau)$ and $R_{yu}(\tau)$ as $N \to \infty$ [Ljung, 1999].

An interpretation of the projection in (3) as $l \to \infty$ is that the product $X\Pi$ approaches an expression containing block-Toeplitz matrices of correlation functions $R_{xu}(\tau)$ and $R_u(\tau)$. Methods for various model types developed from this approach with correlation-function interpretations my be found in Chou and Verhaegen [1997] and its references. Two important distinctions exist between these methods and the preceding method: the block-Toeplitz matrices formed in the asymptotic analysis of (3) will not include information for $\tau < 0$, hence less information is used since $R_{yu}(\tau)$ is not symmetric; and the data matrices of the preceding algorithm remain finite-dimensional as $N \to \infty$, greatly increasing the amount of data on which the identification procedure may be applied.

4. IDENTIFICATION OF INPUT DYNAMICS

An often overlooked step of subspace identification is the identification of the parameters B and D in (1). In the following, we present a memory-efficient method of calculating the regressor matrix that may be used to find least-squares estimates of B, D, and an initial condition $x(0)$. We then extend the method to correlation function estimates.

4.1 Identification from Input-Output Data

With $\hat{A}$ and $\hat{C}$ identified, estimates for B, D and $x(0)$ may be found by minimizing the least-squares error

$$\min_{\hat{B},\hat{D},\hat{x}(0)} ||y - \hat{y}||_2^2, \tag{6}$$

where $\hat{x}(0)$ is an estimate of the initial state $x(0)$,

$$y = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} \in \mathbb{R}^{Nn_y}, \qquad \hat{y} = \begin{bmatrix} \hat{y}(0) \\ \hat{y}(1) \\ \vdots \\ \hat{y}(N-1) \end{bmatrix} \in \mathbb{R}^{Nn_y},$$

and

$$\hat{y}(t) = \hat{C}\hat{A}^t\hat{x}(0) + \sum_{k=0}^{t-1} \hat{C}\hat{A}^{t-k-1}\hat{B}u(k) + \hat{D}u(t). \tag{7}$$

The parameters $\hat{B}$, $\hat{C}$, and $\hat{x}(0)$ are linear in $y - \hat{y}$ and may be written as a right-hand product by use of the Kronecker product, resulting in

$$\begin{aligned} \hat{y}(t) = \hat{C}\hat{A}^t\hat{x}(0) &+ \left(u(t)^T \otimes I_{n_y}\right) \mathrm{vec}(\hat{D}) \\ &+ \left(\sum_{k=0}^{t-1} u(k)^T \otimes \hat{C}\hat{A}^{t-k-1}\right) \mathrm{vec}(\hat{B}). \end{aligned}$$

We then group the unknown parameters into a single vector to form

$$\begin{aligned} \hat{y}(t) &= \begin{bmatrix} \phi_{x_0}(t)^T & \phi_D(t)^T & \phi_B(t)^T \end{bmatrix} \begin{bmatrix} \hat{x}(0) \\ \mathrm{vec}(\hat{D}) \\ \mathrm{vec}(\hat{B}) \end{bmatrix} \\ &= \phi(t)^T\theta. \end{aligned}$$

where

$$\phi_{x_0}(t)^T = \hat{C}\hat{A}^t \in \mathbb{R}^{n_y \times n} \tag{8}$$

$$\phi_D(t)^T = u(t)^T \otimes I_{n_y} \in \mathbb{R}^{n_y \times n_u n_y} \tag{9}$$

$$\phi_B(t)^T = \sum_{k=0}^{t-1} u(k)^T \otimes \hat{C}\hat{A}^{t-k-1} \in \mathbb{R}^{n_y \times n_u n}. \tag{10}$$

Thus the minimization problem (6) is now a linear-least-squares problem of the form

$$\min_{\theta} \frac{1}{N} \sum_{t=0}^{N-1} \left|\left|y(t) - \phi(t)^T\theta\right|\right|_2^2. \tag{11}$$

In practice, a significant computational limitation is the calculation of the regressor $\phi_B^T(t)$. The Kronecker product in (10) results in calculation and summation of $O(N^3/2)$ matrices of size $n_y \times n_u n$. Memory limitations will dramatically limit the maximum N that may be used for computing $\phi(t)^T$, particularly for multivariable data of large dimension. This problem may be compounded when the method is extended to correlation function estimates.

4.2 Identification from Correlation Function Estimates

To extend the least-squares estimation of B, D, and $x(0)$ to correlation function estimates $\hat{R}_u(\tau)$ and $\hat{R}_{yu}(\tau)$ calculated over $\tau \in [\tau_0, \tau_1]$, we replace (6) with

$$e = \min_{\hat{B},\hat{D},\hat{R}_{xu}(\tau_0)} \left|\left|\hat{R}_{yu} - \beta\right|\right|_F \tag{12}$$

where $\hat{R}_{xu}(\tau_0)$ is an estimate of $R_{xu}(\tau_0)$,

$$\hat{R}_{yu} = \begin{bmatrix} \hat{R}_{yu}(\tau_0) \\ \hat{R}_{yu}(\tau_0+1) \\ \vdots \\ \hat{R}_{yu}(\tau_1) \end{bmatrix} \in \mathbb{R}^{Pn_y \times n_u}$$

$$\beta = \begin{bmatrix} \beta(\tau_0) \\ \beta(\tau_0+1) \\ \vdots \\ \beta(\tau_1) \end{bmatrix} \in \mathbb{R}^{Pn_y \times n_u},$$

where $P = \tau_1 - \tau_0 + 1$, and

$$\beta(\tau) = \hat{C}\hat{A}^{\tau}\hat{R}_{xu}(\tau_0) + \sum_{k=0}^{\tau-1} \hat{C}\hat{A}^{\tau-k-1}\hat{B}\hat{R}_u(k) + \hat{D}\hat{R}_u(\tau).$$

Vectorizing $\beta(\tau)$ and separating the unknown parameters $\hat{B}$, $\hat{D}$ and $\hat{R}_{xu}(\tau_0)$ with the Kronecker product results in

$$\begin{aligned} \operatorname{vec}(\beta(\tau)) = &\left(I_{n_u} \otimes \hat{C}\hat{A}^{\tau}\right) \operatorname{vec}(\hat{R}_{xu}(\tau_0)) \\ &+ \left(\hat{R}_u(\tau)^T \otimes I_{n_y}\right) \operatorname{vec}(\hat{D}) \\ &+ \left(\sum_{k=0}^{\tau-1} \hat{R}_u(k)^T \otimes \hat{C}\hat{A}^{\tau-k-1}\right) \operatorname{vec}(\hat{B}), \end{aligned}$$

and we obtain the new regressor

$$\begin{aligned} \operatorname{vec}(\beta(\tau)) &= \begin{bmatrix} \phi_{R_{xu}}(\tau)^T & \phi_D(\tau)^T & \phi_B(\tau)^T \end{bmatrix} \begin{bmatrix} \operatorname{vec}(\hat{R}_{xu}(\tau_0)) \\ \operatorname{vec}(\hat{D}) \\ \operatorname{vec}(\hat{B}) \end{bmatrix} \\ &= \phi(\tau)^T\theta, \end{aligned}$$

where

$$\phi_{R_{xu}}(\tau)^T = I_{n_u} \otimes \hat{C}\hat{A}^{\tau} \in \mathbb{R}^{n_y n_u \times n n_u} \tag{13}$$

$$\phi_D(\tau)^T = \hat{R}_u(\tau)^T \otimes I_{n_y} \in \mathbb{R}^{n_y n_u \times n_y n_u} \tag{14}$$

$$\phi_B(\tau)^T = \sum_{k=0}^{\tau-1} \hat{R}_u(k)^T \otimes \hat{C}\hat{A}^{\tau-k-1} \in \mathbb{R}^{n_y n_u \times n n_u}. \tag{15}$$

Although the computation of correlation function estimates in (5) reduces the number of matrices needed to estimate B from $O(N^3/2)$ to $O\left(P^3/2\right)$, these matrices have grown to size $n_y n_u \times n n_u.$, and the matrices needed to estimate D have grown to size $n_y n_u \times n_y n_u$. The regression problem consequently becomes intractable for high-dimensional multivariable systems without a more efficient means of calculating $\phi_B^T(\tau)$.

4.3 Efficient Computation of the Regressor

We begin by first reformulating the calculation of the regressor for the raw data case and then extend to the method to correlation function estimates . To avoid calculating $\phi_B(t)^T$ from (10) explicitly, we show that $\phi_B(t)^T$ may be calculated more efficiently as a set of state sequences of a dual system.

Theorem 1. Block element (i, j) of the transposed regressor $\phi_B(t)$ in (10) is equivalent to the state sequence of the system

$$\phi_B^{(i,j)}(t+1) = \hat{A}^T\phi_B^{(i,j)}(t) + \hat{C}^T\tilde{u}(i,j,t) \tag{16}$$

computed with the initial condition $\phi_B(0) = 0_n$, in which

$$\tilde{u}(i,j,t) = \begin{bmatrix} 0_{j-1} \\ u_i(t) \\ 0_{n_y-j} \end{bmatrix}.$$

Proof. First, observe that the transpose of (10) may be expanded as

$$\begin{aligned} \phi_B(t) &= \sum_{k=0}^{t-1} \left(u(k)^T \otimes \hat{C}\hat{A}^{t-k-1}\right)^T \\ &= \sum_{k=0}^{t-1} \left[\left(u(k)^T \otimes I_{n_y}\right)\left(I_{n_u} \otimes \hat{C}\hat{A}^{t-k-1}\right)\right]^T \\ &= \sum_{k=0}^{t-1} \left(I_{n_u} \otimes \hat{C}\hat{A}^{t-k-1}\right)^T \left(u(k)^T \otimes I_{n_y}\right)^T \\ &= \sum_{k=0}^{t-1} \left(I_{n_u} \otimes \left(\hat{A}^T\right)^{t-k-1}\hat{C}^T\right)\left(u(k) \otimes I_{n_y}\right). \end{aligned}$$

The first Kronecker product within the summation expands to the block-diagonal matrix

$$\begin{bmatrix} \left(\hat{A}^T\right)^{t-k-1}\hat{C}^T & 0 & \cdots \\ 0 & \left(\hat{A}^T\right)^{t-k-1}\hat{C}^T & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

and multiplication on the right by $u(k) \otimes I_{n_y}$ results in

$$\begin{bmatrix} \left(\hat{A}^T\right)^{t-k-1}\hat{C}^T u_1(k) \\ \left(\hat{A}^T\right)^{t-k-1}\hat{C}^T u_2(k) \\ \vdots \end{bmatrix}, \tag{17}$$

where $u_i(k)$ is the i-th component of the input signal, $i \in [1, n_u]$. Because $u_i(k)$ is scalar, $u_i(k) \otimes I_{n_y} = u_i(k)I_{n_y}$, and we may reincorporate the summation to find

$$\phi_B(t) = \begin{bmatrix} \sum_{k=0}^{t-1} \left(\hat{A}^T\right)^\eta \hat{C}^T \begin{bmatrix} u_1(k) \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \sum_{k=0}^{t-1} \left(\hat{A}^T\right)^\eta \hat{C}^T \begin{bmatrix} 0 \\ u_1(k) \\ 0 \\ \vdots \end{bmatrix} & \cdots \\ \sum_{k=0}^{t-1} \left(\hat{A}^T\right)^\eta \hat{C}^T \begin{bmatrix} u_2(k) \\ 0 \\ 0 \\ \vdots \end{bmatrix} & \sum_{k=0}^{t-1} \left(\hat{A}^T\right)^\eta \hat{C}^T \begin{bmatrix} 0 \\ u_2(k) \\ 0 \\ \vdots \end{bmatrix} & \cdots \\ \vdots & \vdots & \end{bmatrix}$$

where $\eta = t - k - 1$. This is a convolution operation, similar to (7), with $\hat{A}^T$ in place of $\hat{A}$, $\hat{C}^T$ in place of $\hat{B}$, and $\hat{C}$ replaced with I_n. Hence, $\phi_B(t)^T$ may be calculated as separate convolutions of significantly smaller dimension. □

Typically the fastest way to compute ϕ_B^T will be to form the $n_u n_y$ input signals $\tilde{u}(i,j,\tau)$ and to compute state-sequences for each using (16). Efficient routines for computing state-sequences of linear, time-invariant systems of this type are commonly available in numerical software packages, such as MATLAB. The remaining regressors $\phi_{x_0}^T$ and ϕ_D^T are straightforward and far less expensive to compute when using input-output data. We now extend Theorem 1 to correlation function estimates.

Theorem 2. Block element (i,j) of the transposed regressor ϕ_B in (15) may be calculated from state sequences of the system

$$\phi_B^{(i,j)}(\tau+1) = \hat{A}^T \phi_B(\tau) + \hat{C}^T \tilde{u}(i,j,\tau) \qquad (18)$$

with the initial condition $\phi_B(\tau_0) = 0_n$, in which

$$\tilde{u}(i,j,\tau) = \begin{bmatrix} 0_{\gamma-1} \\ \hat{R}_{u_\alpha,\xi_\beta}(\tau) \\ 0_{n_y-\gamma} \end{bmatrix} \quad \begin{array}{l} \alpha = \text{floor}((i-1)/n_u) \\ \beta = \text{floor}((j-1)/n_y) \\ \gamma = \text{mod}(j-1, n_y) \end{array},$$

where $\text{mod}(r,s) : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is r modulo s, and $\text{floor}(r) : \mathbb{R} \to \mathbb{Z}$ is the nearest integer $\leq r$.

Proof. Expand (15) in the same manner as in Theorem 1 to find

$$\phi_B(\tau) = \sum_{k=0}^{\tau-1} \left(I_{n_u} \otimes \left(\hat{A}^T\right)^{\tau-k-1} \hat{C}^T \right) \left(R_u(k) \otimes I_{n_y} \right).$$

The term within the summation – similar to (17) – becomes

$$\begin{bmatrix} \hat{A}^{T^\eta} \hat{C}^T \left(R_{u_1\xi_1}(k) I_{n_y} \right) & \hat{A}^{T^\eta} \hat{C}^T \left(R_{u_1\xi_2}(k) I_{n_y} \right) & \cdots \\ \hat{A}^{T^\eta} \hat{C}^T \left(R_{u_2\xi_1}(k) I_{n_y} \right) & \hat{A}^{T^\eta} \hat{C}^T \left(R_{u_2\xi_2}(k) I_{n_y} \right) & \cdots \\ \vdots & \ddots & \end{bmatrix}$$

where $\eta = t - k - 1$. The above may be interpreted as a convolution as in Theorem 1, and the result follows. □

Additionally, the symmetry of $R_u(\tau)$ may be used to reduce the necessary computation when using correlation function estimates.

5. SIMULATION EXAMPLE

We provide a simulation example that demonstrates both the improved consistency of using correlation function estimates with colored noise and the efficiency of computing the input-to-state regressor using the dual-system approach. Let a linear, time-invariant, discrete-time system be described by the state-space equations (1) with

$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 & 0 \\ -0.04 & 0.3 & 1 & 0 & 0 \\ 0 & 0 & -0.5 & 1 & 0 \\ 0 & 0 & 0 & -0.1 & 1 \\ 0 & 0 & 0 & -0.81 & -0.1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$C = I_5 \qquad D = 0_{5\times 2},$$

so that $n = 5$, $n_u = 2$, and $n_y = 5$. The input $u(t)$ is white noise with unit variance, the initial condition $x(0)$ is a random variable with covariance matrix

$$E[x(0)] = I_n,$$

and the noise-process $v(t) \in \mathbb{R}^{n_y}$ be described by a zero-mean white noise signal $e(t) \in \mathbb{R}^{n_y}$ with correlation function

$$R_e(\tau) = 0.01 I_{n_y} \delta(\tau)$$

that is filtered through a linear, time-invariant system with the fractional description

$$v(t) = \frac{1}{(q-0.8)\,(q^2 - 1.2q + 0.37)} I_{n_y} e(t)$$

where q is the time-shift operator. The estimate $\hat{A}$ was computed three ways: using the MATLAB System Identification Toolbox "n4sid" function [Mathworks, 2009] with MOESP weighting selected, the same function with CVA weighting selected, and from correlation function estimates. For the correlation function estimates, an oblique projection was used to remove the effects of future input, as described in Van Overschee and De Moor [1996b]. The "past" and "future" horizons for the projection of correlation function data were taken from the n4sid results.

Plots of the eigenvalues of $\hat{A}$ estimated from 100 simulations are shown in Fig. 1. Data sequences of $N = 2000$ samples were used to generate estimates for N4SID and correlation function estimates of length $P = 17$. The correlation function estimates are centered around the poles of the true system, while both N4SID weightings are biased towards the poles of the noise-generating system. Plots of the maximum singular values of the identified systems vs. frequency can be seen in Fig. 2.

Not only did the correlation function estimates provide identified systems that more closely match the true systems dynamics, but the dramatic reduction in over matrix sizes also resulted in a roughly 92% decrease in computation time. When the formulas in (13), (14), and (15) were used explicitly, the computations resulted in a memory error, even when MATLAB's "kron" function was reprogrammed to be more efficient.

6. CONCLUSIONS AND FUTURE WORK

We presented a memory-efficient way of estimating input dynamics from correlation functions that. The correlation-function approach was shown to provide consistent estimates even in the case of highly colored noise in a simulation example. Results from correlation function estimates were compared with popular algorithms based on input-output data and were shown to achieve estimates closer to the dynamics of the true system using smaller matrices than were necessary with the input-output data approach.

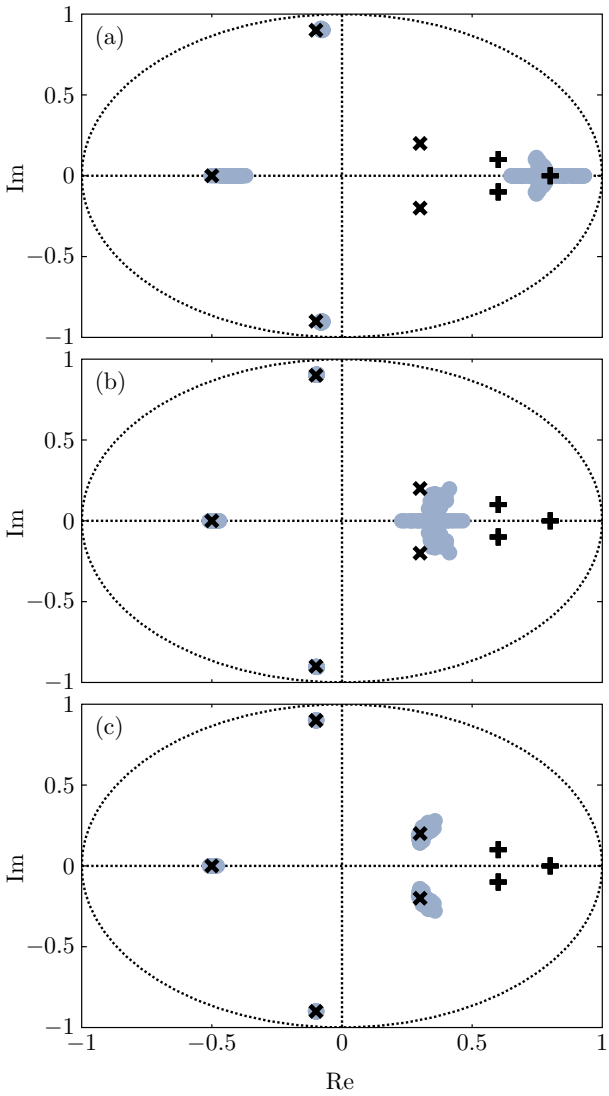

Fig. 1. Plot of pole locations of 100 simulated identifications. 'x' marks the locations of the true system poles, '+' marks the locations of the noise process poles, and the shaded regions are the locations of the identified poles. (a) N4SID with MOESP weighting, (b) N4SID with CVA weighting, (c) correlation function estimates.

Future work in this area might extend the correlation function-based approach to closed-loop identification. If the input and output noise are correlated, cross-correlation functions of the input and output with an external reference signal could be estimated and used to identify unbiased estimates of system dynamics. Also, exploring the application of frequency-domain smoothing techniques on the correlation function estimates could result in control-relevant identification procedures.

REFERENCES

C.T. Chou and M. Verhaegen. Subspace algorithms for the identification of multivariable dynamic Errors-in-Variables models. *Automatica*, 33(10):1857–1869, October 1997.

D. Di Ruscio. A method for the identification of state space models from input and output measurements. *Modeling, Identification and Control: A Norwegian Research Bulletin*, 16(3):129–143, 1995.

T. Katayama. *Subspace Methods for System Identification*. Springer-Verlag, 2005.

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

The Mathworks. Matlab system identification toolbox, 2009.

T. McKelvey, H. Akcay, and L. Ljung. Subspace-based multivariable system identification from frequency response data. *Automatic Control, IEEE Transactions on*, 41(7):960–979, 1996.

D.N. Miller and R.A. de Callafon. Subspace identification from classical realization methods. In *15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

D.N. Miller and R.A. de Callafon. Subspace identication using dynamic invariance in shifted Time-Domain data. In *to be published in Proceedings of the 2010 Conference on Decision and Control*, Atlanta, GA, USA, 2010.

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

P. Van Overschee and B. De Moor. Continuous-time frequency domain subspace system identification. *Signal Processing*, 52(2):179–194, July 1996b.

M. Verhaegen and V. Verdult. *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 1 edition, May 2007.

K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, August 1995.

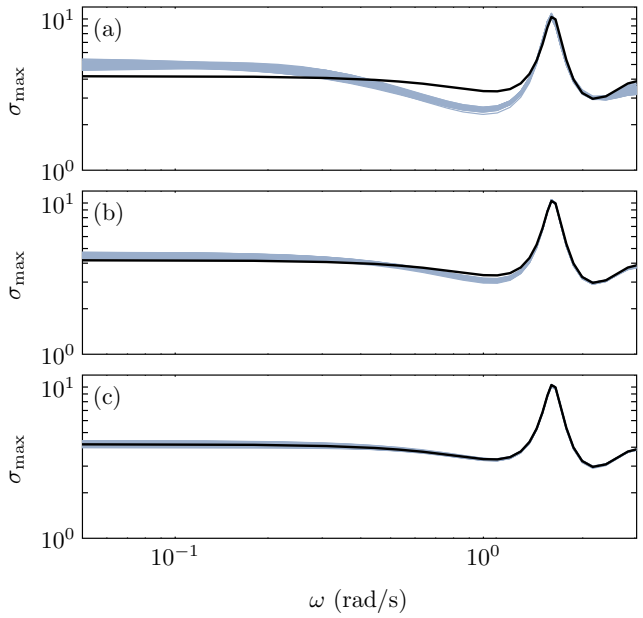

Fig. 2. Maximum singular values and minimum singular values vs. frequency of the true system (black) and 100 simulated identifications (gray). (a) N4SID with MOESP weighting, (b) N4SID with CVA weighting, (c) correlation function estimates.