Identification of multivariable dynamic errors-in-variables system with arbitrary inputs

Erliang Zhang\textsuperscript{a}, Rik Pintelon\textsuperscript{b}

\textsuperscript{a}School of Mechanical Engineering, Zhengzhou University, Science Avenue 100, 450001, Zhengzhou, China
\textsuperscript{b}Department ELEC, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

Abstract

The present work deals with the identification of multivariable linear dynamic systems from noisy input-output observations, where the input signal is arbitrary and the input-output noises are mutually correlated. A frequency domain identification framework is developed, in which the consistent estimator of the multivariable plant model parameters and of the input-output noise covariance matrix is defined as the solution of a set of normal equations and sufficient conditions for the uniqueness of the parameter estimate are established based on the rank property of the matrix of the normal equation. The uncertainty bound of the parameter estimates is constructed and compared with the Cramér-Rao lower bound. The proposed methodology is validated on a simulated multivariable dynamic system.

Key words: Errors-in-variables; multivariable system; arbitrary inputs; frequency domain identification.

1 Introduction

The identification of dynamic errors-in-variables (EIV) systems plays a primal role when the aim is to model the underlying system from input-output observations corrupted by disturbing noise. A rich set of identification methods is available especially for single input and single output (SISO) EIV model, as seen in the survey paper [19] and algorithms reported therein. Efforts are also made to develop EIV identification methods in more general conditions. The identification of SISO dynamic systems with correlated input-output noises has been handled by a bias-eliminating least-squares approach [5], further addressed in an instrumental variable framework which unifies many estimators [21].

Once identifiability is established for a particular EIV application, estimation algorithms can be developed. A few kinds of identification algorithms, which can estimate consistently the multivariable EIV model, have been proposed, such as subspace algorithm [3], Frisch scheme [6], generalized instrumental variable method [20], recursive identification [2], and frequency domain identification [15]. Note that most of these algorithms are developed by exploiting the ergodic property of the inputs.

The objective of the present work is to estimate consistently the multivariable EIV system, excited by arbitrary inputs, from noisy input-output data. A very general EIV setup is considered herein, where the dynamic system is not necessarily causal and minimum phase, the input-output disturbing noises are allowed to be mutually correlated and the assumption of non-white input is removed. This goes beyond the standard EIV assump-
The proposed estimation algorithm formulates the estimator of the multivariable EIV model parameters as a solution of a set of (nonlinear) normal equations, which is a multivariable extension of the frequency-domain estimation algorithm [22]. However, this extension is basically non-trivial due to the increase in size and complexity of multivariable EIV systems, more algorithmic and conceptual difficulties are present when analyzing the consistency and accuracy of the proposed estimator. Moreover, an efficient numerical optimization algorithm is needed for solving large-scale nonlinear normal equations of different kinds of variables.

The identifiability of multivariable dynamic systems in the present EIV setting is still an open question. Another novel contribution of the present work is to provide a sufficient condition that can guarantee the uniqueness of the parameter estimate, which is derived by analyzing the matrix regularity of the normal equations that give consistent parameter estimates.

The rest of the paper is organized as follows. Section 2 formulates the problem of concern. The proposed identification framework is described in Section 3. Section 4 is used to establish sufficient conditions for the uniqueness of the parameter estimate and develop a robust numerical algorithm. Section 5 is devoted to the accuracy analysis. The proposed method is validated on a simulated system in Section 6. Conclusions are drawn in Section 7.

2 Problem formulation

Definition 1 (Discrete Fourier transform) The discrete Fourier transform (DFT) of a time domain signal $x(n)$ is denoted by $X(k)$, and defined as

$$ X(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \exp(-j2\pi k n/N). \quad (1) $$

The multivariable system is linear and time-variant. The following input-output relation holds exactly in the frequency domain when the time domain input-output data are transformed using the DFT,

$$ Y(k) = G(\Omega_k)U_n(k) + T(\Omega_k) + NV(k), \quad (2) $$

$$ U(k) = UP(k) + NU(k), \quad (3) $$

where $U_n(k)$ is the noise-free input spectrum whose true value is $U_n(\omega_k)$, $G(\Omega_k) \in \mathbb{C}^{n_y \times n_u}$ is the transfer function matrix with $n_y$ and $n_u$ the numbers of system inputs and outputs respectively, $T(\Omega_k)$ is the plant transient whose amplitude vanishes as an $O(N^{-1/2})$ with respect to the main contribution [14], $NU(k)$ and $NV(k)$ are the DFTs of the input and output disturbing noises respectively, $\Omega$ is a general frequency variable, $\Omega_k = \exp(-j2\pi k/N)$ for discrete-time systems, $\Omega_k = j2\pi f_k N$ for continuous-time systems with $f_k$ the sampling frequency. The frequency index $k \in K$, whose number of elements, $F$, is increased as $F = \mathcal{O}(N)$. The input-output measurement are collected in $Z$, i.e., $Z^T(k) = [Y^T(k), U^T(k)]$ with the superscript $T$ denoting the transpose operation.

Definition 2 The transfer function model is a rational matrix with a common denominator,

$$ G(\Omega, \theta) = A^{-1}(\Omega, \theta_a)B(\Omega, \theta_b), \quad (4) $$

where $(B(\Omega, \theta_b))_{i,j} = \sum_{r=0}^{n_u} b_{ij}\Omega^r$, and

$$ A(\Omega, \theta_a) = \sum_{r=0}^{n_a} a_{r}\Omega^r \quad \text{with} \quad \begin{cases} a_0 = 1 & \text{if} \ \Omega = z^{-1} \\ a_n = 1 & \text{if} \ \Omega = s \end{cases} \quad (5) $$

Assumption 3 The multivariable system is dynamic, moreover the amplitude of the system (4) is frequency-dependent.

Assumption 4 The polynomial matrix $B(\Omega_0, \theta_b) \neq 0$ if $A(\Omega_0, \theta_a) = 0$.

Assumption 5 The polynomial degrees $n_a$ and $n_b$ are assumed to be known.

The plant transient is a smooth function of the frequency, and has the same poles as $G(\Omega, \theta)$ [14], it is parameterized as

$$ T(\Omega, \theta_a, \theta_c) = A^{-1}(\Omega, \theta_a)C(\Omega, \theta_c), \quad (6) $$

where $C(\Omega, \theta_c) = \sum_{c=0}^{n_c} c_{c}\Omega^c$ with $c_{c} \in \mathbb{R}^{n_u}$, the order $n_c = \max(n_a, n_b) = 1$ for discrete-time systems, while for continuous-time systems, $n_c = \max(n_a, n_b) - 1$.

Assumption 6 The inputs are arbitrary random signals which have the following properties:

(A1) The inputs are mutually uncorrelated.

(A2) The inputs are persistently exciting of sufficiently high order such that $|G(\Omega_k)|U_0(k) \neq 0$, $k \in K$.

The input takes a nonparametric form, which is an $n_u$-dimensional DFT sequence. It can incorporate nonstationary and harmonic components.

Assumption 7 (Noise assumption) (a) The temporal input-output disturbing noises are i.i.d. over the time instants, mutually correlated random variables with covariance matrix $\Sigma$, and uncorrelated with the noise-free input. (b) The temporal input-output noises are jointly normally distributed.
Under Assumption 7, \( N_T(k) \) and \( N_U(k) \) are zero-mean, circular complex normally distributed and independent over the frequency index \( k \) [14]. If Assumption 7(b) is not fulfilled, the statistical property of \( N_T(k) \) and \( N_U(k) \) is asymptotically \((N \to \infty)\) valid. It is remarked that the covariance matrix \( \Sigma \) is the same for the temporal input-output noises and their DFTs given by Definition 1.

**Problem 8** Given the input-output DFTs \( \{Z(k)\}_{k \in K} \), estimate the plant model coefficients \( \theta \) and the input-output noise covariance \( \Sigma \).

### 3 Identification framework

#### 3.1 Gaussian likelihood function

\( Z_p \) is introduced to represent the noise-free counterpart of \( Z \), whose true value is \( Z_0 \). Under Assumption 7, the Gaussian likelihood function of all the unknowns reads,

\[
p(Z|\Phi) \propto |\Sigma|^{-F} \exp \left\{-\frac{1}{2} \sum_{k \in K} \|Z(k) - Z_p(k)\|^2_{\Sigma} \right\}, \tag{7}
\]

where \( |X| = \text{det}(X), \|x\|^2_W = x^H W^{-1} x \) with the superscript \( ^\dagger \) denoting the Hermitian conjugate transpose. \( p(Z|\Phi) \) is subject to the equality constraint

\[
M(k, \vartheta)Z_p(k) - T(\Omega_k, \vartheta) = 0, \tag{8}
\]

where \( M(k, \vartheta) = \left[I_{n_y} - G(\Omega_k, \vartheta)\right] \) with \( I_{n_y} \) an identity matrix, \( \Phi^T = \left[U_p^T, \vartheta^T\right] \) and \( \vartheta^T = \left[\theta^T, \vartheta^T, (\sigma^2)^T\right] \). \( \sigma^2 \) collects the basic variables of the symmetric matrix \( \Sigma \), which are the entries on or above the main diagonal.

A cost function of all the unknowns is defined based on the Gaussian likelihood function (7), \( V_p(\Phi, Z) = -\ln p(Z|\Phi) \),

\[
V_p(\Phi, Z) = F \ln |\Sigma| + \frac{1}{2} \sum_{k \in K} \|Z(k) - Z_p(k)\|^2_{\Sigma^2}, \tag{9}
\]

subject to (8). However, \( \Phi \) cannot be consistently estimated from (9). This is because the unknown variables of the noise-free input \( U_p \) grow with data points \( F \) at the same rate. Hence \( U_p \) should not be jointly estimated with the model parameters \( \vartheta \). Other forms of cost function need to be developed.

#### 3.2 Concentrated likelihood function

The noise-free input is substituted with its maximum in (7), which leads to a concentrated cost function used for maximum likelihood identification [14]. The maximum \( Z_p(k) \) containing \( \hat{U}_p(k) \) is obtained from (7) by applying the method of Lagrange multipliers to take into account the equality constraint (8),

\[
\hat{Z}_p(k) = Z(k) - M^\dagger(k, \vartheta)C_E^{-1}(k, \vartheta) E(k, \vartheta), \tag{10}
\]

where \( C_E(k, \vartheta) = M(k, \vartheta)\Sigma M^\dagger(k, \vartheta) \) and \( E(k, \vartheta) = M(k, \vartheta)Z(k) - T(\Omega_k, \vartheta) \). Substituting \( \hat{Z}_p(k) \) in (7), the cost function is defined as

\[
V_1(\vartheta, Z) = -\ln \left[p\left(Z|\vartheta, \hat{Z}_p\right)\right]. \tag{11}
\]

The basic idea for consistency proof, based on the law of large numbers, is to show that the expected value of the cost function w.r.t. parameters \( \vartheta \) should be zero in their true values \( (\hat{\vartheta}) \). It can be derived that

\[
\frac{\partial V_1(\vartheta)}{\partial \vartheta} \bigg|_{\vartheta = \hat{\vartheta}} = 0, \tag{12}
\]

where \( V_1(\vartheta) = \mathbb{E}[V(\vartheta, Z)] \) with \( \mathbb{E} \) denoting the expectation operation w.r.t. measurements. However, the derivative of \( V_1(\vartheta) \) w.r.t. the noise covariance,

\[
\frac{\partial V_1(\vartheta)}{\partial \sigma^2} \bigg|_{\vartheta = \hat{\vartheta}} \neq 0. \tag{13}
\]

It is concluded that \( \vartheta \) and \( \sigma^2 \) cannot be estimated consistently from (11) only. The proof of (12) and of (13) is given in Appendix B.

#### 3.3 Marginal likelihood function

Considering \( U_p \) as a nuisance parameter, the likelihood function is formulated only in terms of \( \vartheta \) by marginalizing over \( U_p \). Assuming that there exists a prior probability distribution of \( U_p(k) \), \( p(U_p(k)|\vartheta) \), which is uniformly distributed and independent over the frequency index \( k \), then

\[
p(Z(k)|\vartheta) = \int p(Z(k)|U_p(k), \vartheta) p(U_p(k)|\vartheta) dU_p(k) \propto \int p(Z(k)|U_p(k), \vartheta) dU_p(k). \tag{14}
\]

See Appendix C for the details of the derivation of \( p(Z(k)|\vartheta) \). The second cost function is thus defined as \( V_2(\vartheta, Z) = -\ln \left[p(Z|\vartheta)\right] \) with \( p(Z|\vartheta) = \prod_{k \in K} p(Z(k)|\vartheta) \).

\[
V_2(\vartheta, Z) = \sum_{k \in K} \ln |C_E(k, \vartheta)| + \|E(k, \vartheta)\|^2_{C_E(k, \vartheta)}. \tag{15}
\]
Analogously, the derivatives of the expected value of the cost function, \( V_2(\theta) = \mathbb{E}[V_2(\theta, Z)] \), are derived for consistency examination,

\[
\frac{\partial V_2(\theta)}{\partial \sigma^2} \bigg|_{\theta_0} = 0, \quad (16)
\]
\[
\frac{\partial V_2(\theta)}{\partial \theta} \bigg|_{\theta_0} \neq 0. \quad (17)
\]

It is concluded that \( \theta \) and \( \sigma^2 \) cannot be estimated consistently from (15) only. The proof of (16) and of (17) is given in Appendix D.

**Remark 9** \( p(Z|\theta) \), from which \( V_2(\theta, Z) \) is derived, is actually the probability density function that describes the statistical properties of the residual error of the equation \( E(k, \theta) \) when \( \theta = \theta_0 \).

### 3.4 Proposed estimator

As suggested by (12) and (16), it is intuitive to solve the following nonlinear normal equations

\[
f(\theta) = \begin{bmatrix}
\frac{\partial V_1(\theta, Z)}{\partial \theta} \\
\frac{\partial V_2(\theta, Z)}{\partial \theta} \\
\frac{\partial V_2(\theta, Z)}{\partial \sigma^2}
\end{bmatrix}
\bigg|_{\theta_0} = 0,
\]

for the identification of multivariable EIV model, where

\[
\hat{\theta}^T = [\theta^T, \theta_c^T].
\]

The solution of (18) defines an estimator for the multivariable EIV model. The properties of the proposed estimator, such as uniqueness and consistency, are studied in the next section, followed by a numerical algorithm that solves (18).

### 4 Parameter estimation

#### 4.1 Local uniqueness

The identifiability analysis is still a challenging issue for general EIV systems defined by Assumptions 3, 6 and 7. As a matter of fact, solving (18) returns the consistent estimate of the EIV model, therefore it is sufficient for EIV identification if conditions to obtain the solution uniqueness of (18) can be established. In this work, they are derived based on the regularity of the Jacobian matrix of (18),

\[
H(\theta, Z) =
\begin{bmatrix}
\frac{\partial^2 V_1(\theta, Z)}{\partial \theta^2} & \frac{\partial^2 V_1(\theta, Z)}{\partial \theta \partial \sigma^2} \\
\frac{\partial^2 V_2(\theta, Z)}{\partial \theta^2} & \frac{\partial^2 V_2(\theta, Z)}{\partial \theta \partial \sigma^2} \\
\frac{\partial^2 V_2(\theta, Z)}{\partial \sigma^2 \partial \theta} & \frac{\partial^2 V_2(\theta, Z)}{\partial (\sigma^2)^2}
\end{bmatrix}. \quad (19)
\]

Since \( \frac{1}{n}H(\theta, Z) \) converges to the expected value \( \frac{1}{n}H(\theta) \) at a rate of \( O(F^{-1/2}) \), \( H(\theta) \) is used instead to derive conditions that obtain the solution uniqueness. In the true parameter values \( \theta_0 \),

\[
H(\theta_0) = \begin{bmatrix} M_1 & M_4 \\ 0 & M_3 \end{bmatrix},
\]

where \( M_1 = 2\text{Re}(J_1^1J_1^1) \bigg|_{\theta_0} \), \( M_3 = \text{Re}(J_3^1J_3^1) \bigg|_{\theta_0} \) and

\[
(M_4)_{i,j} = \text{Re} \left\{ \sum_{k \in \Omega} \left[ \frac{\partial C_E(k, \theta)}{\partial \theta[i]} C_E^{-1}(k, \theta) \frac{\partial C_E(k, \theta)}{\partial \sigma[i]} \right. \\
\left. \times C_E^{-1}(k, \theta) - \frac{\partial^2 C_E(k, \theta)}{\partial \theta[i] \partial \sigma[i]} C_E^{-1}(k, \theta) \right] \right\} \bigg|_{\theta_0}
\]

\[
(J_1)_{i,j} = \left\{ \left[ C_E^{-\frac{1}{2}}(k, \theta) \right]^\dagger \otimes C_E^{-\frac{1}{2}}(k, \theta) \right\} \text{vec} \left[ \frac{\partial C_E(k, \theta)}{\partial \sigma[i]} \right] \\
(J_2)_{i,j} = C_E^{-\frac{1}{2}}(k, \theta) \frac{\partial G_{ij}(\Omega_k, \theta)}{\partial \theta[i]} U_0(k),
\]

where \( C_E^{-1}(k, \theta) = \left[ C_E^{-\frac{1}{2}}(k, \theta) \right]^\dagger C_E^{-\frac{1}{2}}(k, \theta) \), \( \text{Re}(\bullet) \) denotes the real part of a complex variable, \( \text{tr}(\bullet) \) represents the trace of a square matrix and \( \text{vec}(\bullet) \) stacks the columns of a matrix on top of each other. The derivation of (20) is given in Appendix E.

**Assumption 10** The inputs are persistent such that \( \zeta_a \) and \( \zeta_b^{ij} \) have at least \((n_a+1)\) and \((n_b+1)\) nonzero values, respectively,

\[
\zeta_a(k) = A^{-1}(k, \theta) C_E^{-\frac{1}{2}}(k, \theta) G(\Omega_k, \theta) U_0(k)
\]
\[
\zeta_b^{ij}(k) = A^{-1}(k, \theta) C_E^{-\frac{1}{2}}(k, \theta) e_{ij} U_0(k)
\]

where \( i = 1, \cdots, n_y, j = 1, \cdots, n_u, e_{ij} \in \mathbb{R}^{n_y \times n_u} \) is a matrix with unique nonzero identity at the \( i \)-th row and the \( j \)-th column.

Note that \( \zeta_a \) and \( \zeta_b \) are the vector coefficients of the entries of \( J_1 \).

**Theorem 11** Under Assumptions 3 - 6 and 10, the matrix \( H(\theta_0) \) has full rank.

**Proof** It is sufficient to demonstrate that \( M_1 \) and \( M_3 \) are both regular. This boils down to the rank property of \( J_1 \) and \( J_3 \) as \( M_1 \) and \( M_3 \) are Gramian matrices. See details in Appendix F.

**Remark 12** \( M_1 \) is the Hessian of the expected value of the cost function \( V_1(\theta) \) when the input-output noise covariance matrix \( \Sigma \) is given, which is guaranteed to be
well-posed by using the persistent excitation properties in Assumptions 6 and 10. \( M_3 \) is the Hessian of \( V_2(\theta) \) when the plant model parameters \( \theta \) are known, which is assumed to be invertible by exploiting the dynamic property (e.g., frequency-dependent amplitude) of the system in Assumption 3.

Corollary 13 Under Theorem 11, the solution of (18) is uniquely determined in the vicinity of \( \theta_0 \) as \( F \to \infty \).

Proof There exists a positive constant \( \epsilon \) such that
\[
H(\theta) = H(\theta_0) + \Delta(\epsilon)
\]
for \( \| \theta - \theta_0 \|_2 \leq \epsilon \). The rank of the matrix \( H(\theta) \) cannot be reduced by the small perturbation \( \Delta(\epsilon) \) provided that the \( L_2 \) norm of \( \Delta(\epsilon) \) is less than the minimal nonzero singular value of \( H(\theta_0) \) [10].

4.2 Strong consistency

Assumption 14 The input-output disturbing noises \( N_Y(k) \) and \( N_U(k) \) are mutually correlated, i.i.d. random variables with finite moments of order 4.

Theorem 15 (Strong consistency) Under Corollary 13 and Assumption 14, the solution of (18) is a strong consistent estimate of the true values \( \theta_0 \) and \( \sigma^2_0 \).

Proof The proposed estimator is defined as the solution of the vector valued function (18). The scaled entries of \( f(\theta), \frac{1}{2} \frac{\partial v_1(\theta, Z)}{\partial \theta} \) and \( \frac{1}{2} \frac{\partial v_2(\theta, Z)}{\partial \sigma} \), converge strongly to their expected values according to the strong law of large numbers at the rate of \( O\left(F^{-1/2+\epsilon}\right) \) with \( \epsilon \) an arbitrary small positive value (see Theorem 4.3.1 of [9] by considering Assumption 14). These expected values equal zero in the true values of plant model parameters and noise covariances, as shown in (12) and (16). Moreover, the solution \( \hat{\theta} \) and \( \tilde{\sigma}^2 \) of (18) is uniquely determined under Corollary 13, thus they converge with probability 1 to the true values \( \theta_0 \) and \( \sigma^2_0 \) (see [18]).

4.3 Numerical algorithm

Often, the nonlinear normal equation (18) can be solved by a line search optimization algorithm or model trust region approach [4]. The present work implements a Powell’s dogleg strategy which works by combining the Newton and steepest descent directions.

\[
h_N = -H^{-1}(\theta, Z)f(\theta), \quad h_{sd} = -\alpha H'(\theta, Z)f(\theta),
\]
where \( h_N \) is the Newton step and \( h_{sd} \) is the step along the steepest descent direction with \( \alpha \) the step length. Both candidates for the step to take from the current point are combined by the Powell’s strategy to choose an appropriate step controlled by the radius of the trust region. This guarantees the monotonic decrease of the cost function, and enlarges the globally convergent region (see [16] for the details). The evaluation of \( H(\theta, Z) \) is computationally involved and it might be numerically ill-conditioned, the computation of \( h_N \) is delicate as a result. Here \( H(\theta, Z) \) is approximated by neglecting second-order derivative information,

\[
H(\theta, Z) \approx 2\text{Re}\left(J^*J\right),
\]
where

\[
J = \begin{bmatrix}
\frac{\partial \delta_1}{\partial \theta} & \frac{\partial \delta_1}{\partial \sigma^2}
\frac{0}{\partial \delta_2}
\end{bmatrix},
\]

By rewriting \( f(\theta) = 2\text{Re}(J^*\delta) \) with \( \delta^T = \begin{bmatrix} \delta_1, (r\delta_2)^T \end{bmatrix} \), \( h_N \) can be reliably computed by solving an over-determined set of equations [14]. The constraint \( \| \theta \|_2 = 1 \) is used instead of the particular parameter constraint (see Definition 2) in order to avoid any ill-conditioned problem.

High quality starting values are necessary, they can be initiated for instance using a subspace algorithm [3] or Frisch scheme [8].

5 Uncertainty bound

5.1 Covariance matrix of parameter estimates

Assumption 16 The input-output disturbing noises \( N_Y(k) \) and \( N_U(k) \) are mutually correlated, i.i.d. random variables with existing moments of order \( 4 + \xi \) (\( \xi > 0 \)).

Theorem 17 Under Assumption 16, \( \hat{\theta} - \theta \) and \( \tilde{\sigma}^2 - \sigma^2 \) are asymptotically zero-mean Gaussian variables with the covariance matrix

\[
\text{Cov} \left( \hat{\theta} \right) = V^{-1}(\theta_0)Q(\theta_0) \left[ V^{-1}(\theta_0) \right]^T,
\]
where

\[
V(\theta_0) = \begin{bmatrix}
\frac{\partial^2 V_1(\theta, \sigma^2_0)}{\partial \theta^2} & \frac{\partial^2 V_1(\theta, \sigma^2_0)}{\partial \theta \partial \sigma^2}
\frac{\partial^2 V_2(\theta, \sigma)}{\partial \sigma^2} & \frac{\partial \theta \partial \sigma^2}{\partial \sigma^2}
\end{bmatrix}_{\theta_0},
\]
and

\[
Q(\theta_0) = E \left[ V^T(\theta_0, Z)V'(\theta_0, Z) \right] \text{ with }
\]

\[
V'(\theta_0, Z) = \begin{bmatrix}
\frac{\partial V_1(\theta, \sigma^2_0, Z)}{\partial \theta}, \frac{\partial V_2(\theta_0, \sigma^2, Z)}{\partial \sigma^2}
\end{bmatrix}_{\theta_0}.
\]
Proof Follow the same lines of the proof of Theorem 11 [22].

Under Assumption 7, \(Q(\vartheta_0)\) and \(V''(\vartheta_0)\) can be elaborated explicitly as

\[
Q(\vartheta_0) = \begin{bmatrix} M_1 + M_2 & 0 \\ 0 & M_3 \end{bmatrix}, \quad V''(\vartheta_0) = \begin{bmatrix} M_1 & M_4 \\ 0 & M_3 \end{bmatrix}, \tag{27}
\]

where \(M_1, M_3\) and \(M_4\) are given in (20), and

\[
(M_2)_{i,j} = 2 \sum_{k \in \mathcal{K}} \text{Re} \left\{ \text{tr} \left[ \frac{\partial M^T(k, \vartheta) C_{\vartheta}^{-1}(k, \vartheta) \partial M(k, \vartheta)}{\partial \vartheta^j} \right] \right\} \Sigma_{\vartheta}^{1/2} \Sigma_{\vartheta}^{-1/2} \Sigma_{\vartheta}^{1/2} \Sigma_{\vartheta}^{-1/2} \right\}_{\vartheta_0}
\]

The derivation of \(Q(\vartheta_0)\) is reported in Appendix G, and \(V''(\vartheta_0)\) is obtained in the same way as \(H(\vartheta_0)\). Inserting (27) back into (25), the full covariance matrix of all the parameter estimates is computed, from which the covariance matrix of the plant model parameter estimates and the one of the estimated covariances are extracted,

\[
\text{Cov}(\hat{\vartheta}) = M_1^{-1}(M_1 + M_2 + M_3 M_3^{-1} M_4^T) M_1^{-1}, \quad \text{Cov}(\hat{\sigma}^2) = M_3^{-1}. \tag{28}
\]

\[\text{Cov}(\hat{\vartheta}) = M_1^{-1}(M_1 + M_2 + M_3 M_3^{-1} M_4^T) M_1^{-1}, \quad \text{Cov}(\hat{\sigma}^2) = M_3^{-1}. \tag{29}\]

\subsection{5.3 Uncertainty bound of transfer function matrix}

The asymptotic normality property of the estimate \(\hat{\vartheta}\) enables to compute the variance of the transfer function estimate based on a first-order Taylor series expansion. e.g., for the transfer function \(G_{[i,j]}\) relating the \(i\)-th output and the \(j\)-th input,

\[
\sigma^2_{\hat{G}_{[i,j]}(\Omega, \vartheta)} \approx \left. \text{Cov}(\hat{\vartheta}) \right| \left[ G_{[i,j]}(\Omega, \vartheta) \right]_{\vartheta_0} \tag{30}
\]

where \(\vartheta\) is the derivative of \(G_{[i,j]}(\Omega, \vartheta)\) w.r.t. \(\vartheta\).

\section{6 Simulated example}

The proposed identification approach is illustrated on a second-order discrete-time multivariable dynamic system, \(n_u = n_y = 2\). The true plant model is \(A(z^{-1}) = 1 - 1.7599 z^{-1} + 0.9571 z^{-2}\), and

\[
B(z^{-1}) = \begin{bmatrix} 0.1900 + 0.1783 z^{-1}, 0.0838 + 0.1219 z^{-1} \\ 0.0838 + 0.1219 z^{-1}, 0.0972 + 0.0037 z^{-1} \end{bmatrix}
\]

One input exciting the system is a \textit{band-limited} white noise, the other is a mixed sequence of a \textit{band-limited} white noise and a sinusoidal component \(\sin(\pi f_s t / 6)\). The sampling frequency \(f_s = 1024\) Hz, the input and output signals are simulated for a duration of 64 seconds. The noise-free input-output signals are perturbed by normally distributed and mutually correlated noises. The correlation coefficient between any two noise components is set to 0.3. The frequency data below 170.3438 Hz (except the DC component) are used for parameter estimation, \(F = 10903\). The reciprocity property of the plant model is considered for system identification.

Monte Carlo simulations are conducted to validate the statistical properties of the estimators. A series of signal-to-noise ratios (SNRs) is considered, which varies from 8 dB to 20 dB. A scalar measure \(\gamma_{\text{RMSE}}\), so-called normalized root mean square error, is defined as,

\[
\gamma_{\text{RMSE}} = \frac{1}{\|\hat{\vartheta}_0\|_2} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \|\hat{\vartheta}^{[i]} - \vartheta_0\|^2_2}, \tag{33}
\]

where \(n = 100\), \(\hat{\vartheta}^{[i]}\) is the \(i\)-th Monte Carlo result. \(\gamma_{\text{RMSE}}\) is also predicted based on the covariance matrix (28) or CRLB matrix (31). Figure 1 shows for the plant model parameter estimates that the predicted uncertainty agrees with the one empirically estimated over
Based on this, the non-parametric modeling of the noise-free input introduces an additional term $M^{-1}_2 (M_2 + M_4 M_3^{-1} M_2^{-1}) M^{-1}_2$ in (28) in comparison with (31), which influences the efficiency of the plant model parameter estimate, however it fortunately diminishes with the increase of SNR. The input-output noise covariance estimates totally lose their efficiency as (29) and (32) are in distinct form, the value of (32) is of the order $O(\sigma^4/F)$ and independent of the plant model.
which concludes (12) by using (A.1). Similarly, using (A.2) and (B.1),

\[
\frac{\partial V_1(\theta)}{\partial \sigma^2_{[i]}} = \sum_{k \in K} \left\{ \text{tr} \left[ \Sigma^{-1}_{kk} \frac{\partial \Sigma}{\partial \sigma^2_{[i]}} \right] \phi_k + \text{tr} \left[ \frac{\partial C_E(k, \theta)}{\partial \sigma^2_{[i]}} \phi_k \right] C_E(k, \vartheta_0) \right\}
\]

it is concluded that (13) holds using (A.1).

C Marginal likelihood function

Factorizing the covariance matrix \(\Sigma\) as

\[
\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yu} \\ \Sigma_{yu}^\dagger & \Sigma_{uu} \end{bmatrix}, \quad \text{(C.1)}
\]

one obtains from (7)

\[
p(U(k) | U_p(k), \theta) = \mathcal{N}(U_p(k), \Sigma_{uu}), \quad \text{(C.2)}
\]

\[
p(Y(k) | U(k), U_p(k), \theta) = \mathcal{N}(Y(k), \Sigma) \quad \text{(C.3)}
\]

with \(Y(k) = G(k, \theta)U_p(k) + \Sigma_{yu} \Sigma_{uu}^{-1}[U(k) - U_p(k)] - T(\Omega_k)\) and \(\Sigma = \Sigma_{yy} - \Sigma_{yu} \Sigma_{uu}^{-1} \Sigma_{yu}^\dagger\). Introducing \(\tilde{G}(\Omega_k, \theta) = G(\Omega_k, \theta) - \Sigma_{yu} \Sigma_{uu}^{-1}\) and \(\tilde{Y}(k) = Y(k) - \Sigma_{yu} \Sigma_{uu}^{-1}U(k) - T(\Omega_k, \theta)\), then (14) is written as

\[
\int p(Y(k) | U(k), U_p(k), \theta) p(U(k) | U_p(k), \theta) \, dU_p(k)
\]

\[
\propto \int \exp \left\{ -\left\| U_p(k) - U_p(k) \right\|_{\Sigma(\theta)}^2 \right\} \, dU_p(k)
\]

\[
\times \exp \left\{ \tilde{U}_p(k) \tilde{\Sigma}^{-1}(k) \tilde{Y}(k) + \tilde{Y}(k) \tilde{\Sigma}^{-1}(k) \tilde{U}_p(k) + \tilde{Y}(k) \tilde{\Sigma}^{-1}(k) \tilde{Y}(k) + \tilde{U}_p(k) \tilde{\Sigma}^{-1}(k) \tilde{U}_p(k) \right\}
\]

\[
\text{(C.4)}
\]

where \(\alpha = (|\Sigma_{uu}| |\Sigma|)^{-1}\),

\[
\tilde{\Sigma}(k) = \left[ \tilde{G}(\Omega_k, \theta) \Sigma(k) \tilde{G}(\Omega_k, \theta) + \Sigma_{uu}^{-1} \right]^{-1}, \quad \text{(C.5)}
\]

\[
\tilde{U}_p(k) = \tilde{\Sigma}(k) \left[ \Sigma_u^{-1} U(k) + \tilde{G}(\Omega_k, \theta) \Sigma_{uu}^{-1} \tilde{Y}(k) \right], \quad \text{(C.6)}
\]

It holds that \(\int \exp \left\{ -\left\| U_p(k) - \tilde{U}_p(k) \right\|_{\tilde{\Sigma}(\theta)}^2 \right\} \, dU_p(k) = \pi^n |\tilde{\Sigma}|\), one can obtain \(\alpha |\tilde{\Sigma}| = |C_E(k, \vartheta)|^{-1}\) by using

\[
|\tilde{\Sigma}^{-1}(k)| = \left| \Sigma_{uu}^{-1} \right| |\Sigma_{uu} + \tilde{G}(\Omega_k, \theta) \Sigma_{uu} \tilde{G}(\Omega_k, \theta) \Sigma_{uu}^{-1}|,
\]

\[
C_E(k, \vartheta) = \Sigma + \tilde{G}(\Omega_k, \theta) \Sigma_{uu} \tilde{G}(\Omega_k, \theta), \quad \text{(C.7)}
\]

where (C.7) is obtained by using the definitions of \(G(\Omega_k, \theta)\) and \(Z\). The exponent in (C.4) is simplified as \(|E(k, \vartheta)|^2_{C_E(k, \vartheta)}\) by using the definitions of \(Y(k)\) and of \(\tilde{\Sigma}\), and

\[
\tilde{\Sigma} = \Sigma_{uu} - \Sigma_{uu} \Sigma_{uu}^{-1} G(\Omega_k) C_E^{-1}(k, \theta) G(\Omega_k) \Sigma_{uu}. \quad \text{(C.8)}
\]

(C.8) is derived by applying the matrix inversion lemma to (C.5) and by using (C.7). Proof of the cost function \(V_2\) is done.

D Proof of (16) and of (17)

The expected value of the cost function \(V_2(\vartheta, Z)\) is

\[
V_2(\vartheta) = \sum_{k \in K} \ln \left| C_E(k, \vartheta) \right| + \text{tr} \left\{ C_E^{-1}(k, \vartheta) \times \left[ |M(k, \vartheta) Z_0(k)|^2 + C_E(k, \vartheta, \sigma_0^2) \right] \right\}. \quad \text{(D.1)}
\]
Differentiating $V_2(\vartheta)$ w.r.t. $\sigma^2_{[i]}$, and using (B.1) and (A.2), it follows that

$$\frac{\partial V_2(\vartheta)}{\partial \sigma^2_{[i]}}|_{\phi_0} = \sum_{k \in K} \left[ C_{E}^{-1}(k, \vartheta_0) \frac{\partial C_{E}(k, \vartheta)}{\partial \sigma^2_{[i]}} \right]_{\phi_0} + \text{tr} \left[ \frac{\partial C_{E}^{-1}(k, \vartheta)}{\partial \sigma^2_{[i]}} \right]_{\phi_0} C_{E}(k, \vartheta_0),$$

which concludes (16) by considering (A.1). Using (B.1), the expression of $J$ is concluded that

$$\frac{\partial V_2(\vartheta)}{\partial \theta_{[i]}}|_{\phi_0} = \sum_{k \in K} \left[ C_{E}^{-1}(k, \vartheta_0) \frac{\partial C_{E}(k, \vartheta)}{\partial \theta_{[i]}} \right]_{\phi_0} + \text{tr} \left[ \frac{\partial C_{E}^{-1}(k, \vartheta)}{\partial \theta_{[i]}} \right]_{\phi_0} C_{E}(k, \vartheta_0) + C_{E}^{-1}(k, \vartheta_0) \frac{\partial C_{E}(k, \vartheta, \sigma^2_{[i]})}{\partial \theta_{[i]}}|_{\phi_0}.$$

Therefore, (17) is proved by using (A.1).

**E Elements of $H(\theta_0)$**

Using (A.1), the expression of $V_1(\vartheta)$ in (B.2) and

$$\text{tr} \left[ \frac{\partial^2 C_{E}(k, \vartheta)}{\partial x \partial y} \right]_{\phi_0} C_{E}(k, \vartheta) = -\text{tr} \left[ C_{E}^{-1}(k, \vartheta) \frac{\partial C_{E}(k, \vartheta)}{\partial x \partial y} \right]_{\phi_0} - 2 C_{E}^{-1}(k, \vartheta) \frac{\partial C_{E}(k, \vartheta)}{\partial x} C_{E}^{-1}(k, \vartheta) \frac{\partial C_{E}(k, \vartheta)}{\partial y}$$

it is concluded that

$$\frac{\partial^2 V_1(\vartheta)}{\partial \theta_{[i]} \partial \theta_{[j]}}|_{\phi_0} = (M_1)_{[i,j]}, \quad (E.2)$$

$$\frac{\partial^2 V_2(\vartheta)}{\partial \sigma^2_{[i]} \partial \theta_{[j]}}|_{\phi_0} = (M_4)_{[i,j]}. \quad (E.3)$$

Utilizing (A.1), (A.2) and (E.1), the second-order derivative of $V_2(\vartheta)$ w.r.t. $\sigma^2$ is derived and evaluated in the true values $\vartheta_0$. This concludes $M_3$. Similarly, it can be easily verified that $\forall i,j$

$$\frac{\partial^2 V_2(\vartheta)}{\partial \sigma^2_{[i]} \partial \theta_{[j]}}|_{\phi_0} = 0. \quad (E.4)$$

**F Proof of full rank property of $J$ and $J_3$**

1) $J_1$ is partitioned as $J_1 = [J_{1;\theta_a}, J_{1;\theta_b}]$ w.r.t. $\theta_a$ and $\theta_b$, $J_{1;\theta_a} = [J_{1;\theta_a}^1, \ldots, J_{1;\theta_a}^v]$ and $\theta_b^j = [b_{0j}^1, \ldots, b_{n_bj}^j]^T$. For any $X = [V, W]$, it holds that

$$\text{rank}(X) = \text{rank}(V) + \text{rank}(W) - \text{dim}[\mathcal{R}(V) \cap \mathcal{R}(W)]$$

with $\mathcal{R}$ denoting the range of a matrix. In what follows, we show that all the submatrices of $J_1$ have full rank and the spaces spanned by their columns are disjoint.

Firstly, the rational basis $\Psi$ of the transfer function, $\Psi_{[k,j]} = [1, \Omega_k, \Omega_k^2, \ldots]$, is a Vandermonde matrix which has full column rank when $F \geq \max\{n_a, n_b\} + 1$. $J_{1;\theta_a}$ and $\{J_{1;\theta_b}^j\}$ can be then proven to have full column rank under Assumption 10 by following the same lines as Example 4.3.4 of [10].

Secondly, $\forall i$, prove $J_{1;\theta_b}$ to be of full rank, i.e., the ranges of $\{J_{1;\theta_b}^j\}_{j=1}^{n_b}$ are disjoint, we use the proof by contradiction. Assume that there exists a vector in the range of $J_{1;\theta_b}^j$, which also belongs to the subspace formed by the ranges of $\{J_{1;\theta_b}^j\}_{j \neq p}$, and utilizing (21), we get

$$\left( \sum_{q=1,q \neq p}^{n_b} x_q \Psi[k,j] \right) \sum_{l=1}^{n_a+1} y_l \Psi[k,l] = 0,$$

where $\forall k \in K, \{x_q, y_l\}$ are the real coefficients. This contradicts the condition (A1) of Assumption 6.

Thirdly, $\forall p, q$, the matrices $\Xi_p$ and $\Xi_q$

$$\Xi_p(k) = [e_{p1}U_0(k), \ldots, e_{pm}U_0(k)], \quad (F.2)$$

$$\Xi_q(k) = [e_{q1}U_0(k), \ldots, e_{qm}U_0(k)] \quad (F.3)$$

are by construction linearly independent if $p \neq q$. It is easily concluded that the set of block matrices, $\{J_{1;\theta_b}^j\}_{j=1}^{n_b}$, are linearly independent. Therefore, $J_{1;\theta_b}$ is proven to be of full rank.

Fourthly, prove the ranges of $J_{1;\theta_a}$ and of $J_{1;\theta_b}$ to be disjoint. Assume that there exists a vector in the range of $J_{1;\theta_a}$, which also belongs to the space spanned by the columns of $J_{1;\theta_b}, \forall k \in K$,

$$\left[ G(\Omega_k) - \tilde{G}(\Omega_k) \right] U_0(k) = 0, \quad (F.4)$$

where

$$\tilde{G}(\Omega_k) = \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} e_{ij} \sum_{l=1}^{n_a+1} x_{ijl} \Psi[k,l] \sum_{l=2}^{n_a+1} y_l \Psi[k,l]$$

in the case of discrete-time systems (considering the coefficient constraint (5)). Evidently, $G(\Omega_k)$ has distinct
poles from \( \tilde{G}(\Omega_k) \). Around the poles \( \{ p_i \} \) of \( G(\Omega_k) \), (F.4) is simplified as
\[
\lim_{\Omega_k \to \{ p_i \}} G(\Omega_k) U_0(k) \approx 0 \quad (F.5)
\]
This is in contradiction with the condition (A2) of Assumption 6. As a result, the ranges of \( J_{1\theta_0} \) and of \( J_{1\theta_0} \) are disjoint. Proof of the full rank property for \( J_3 \) is done.

2) \( \partial \Sigma / \partial \sigma^2_i \) is denoted by \( Y_i \), the set of symmetric matrices \( \{ Y_i \} \) are by construction linearly independent as anyone cannot be linearly approximated by the others.
\[
\sum_i x_i \frac{\partial C_E(k, \vartheta)}{\partial \sigma^2_i} = [M(k, \vartheta)A][M(k, \vartheta)A]^\dagger = 0 \quad (F.6)
\]
with \( \forall k \in K, \sum_i x_i Y_i = \Lambda \Lambda^\dagger \), (F.6) implies that \( M(k, \vartheta)A = 0 \) for all frequencies, i.e., \( M(k, \vartheta) \) must be a constant matrix. This is in contradiction with the frequency dependent property of \( M(k, \vartheta) \) by Assumption 3. Then \( J_3 \) has full column rank by considering the fact that \( C_E^{-2}(k, \vartheta) \otimes C_E^{-2}(k, \vartheta) \) is invertible for all frequencies.

G  Elements of \( Q(\vartheta_0) \) in (27)

\( Q(\vartheta_0) \) consists of three basic block matrices, which are defined as
\[
(Q_1)_{[i,j]} = E \left[ \frac{\partial V_1(\vartheta, \sigma^2_i, Z)}{\partial \theta_{[i]}} |_{\vartheta_0} \right] \\
(Q_2)_{[i,j]} = E \left[ \frac{\partial V_1(\vartheta, \sigma^2_i, Z)}{\partial \vartheta_{[i]}} |_{\sigma^2_0} \right] \\
(Q_3)_{[i,j]} = E \left[ \frac{\partial V_2(\theta_0, \sigma^2_i, Z)}{\partial \sigma^2_{[j]}} |_{\sigma^2_0} \right]
\]
By introducing the notations,
\[
\Pi_i(k) = \left. \frac{\partial [M^\dagger(k, \vartheta) C_E^{-1}(k, \vartheta)M(k, \vartheta)]}{\partial \theta_{[i]}} \right|_{\vartheta_0}, \quad (G.1)
\]
\[
W_i(k) = \left. \frac{\partial C_E^{-1}(k, \vartheta)M(k, \vartheta)}{\partial \sigma^2_{[i]}} \right|_{\vartheta_0}, \quad (G.2)
\]
\[
\Lambda_i(k) = \left. \frac{\partial C_E^{-1}(k, \vartheta)M(k, \vartheta)}{\partial \theta_{[i]}} \right|_{\vartheta_0}, \quad (G.3)
\]
and \( N_Z^T(k) = [N_Y^T(k), N_U^T(k)] \).
\[
\frac{\partial V_1(\vartheta, \sigma^2_i, Z)}{\partial \theta_{[i]}} |_{\theta_0} = \sum_{k \in K} N_Z^T(k) \Pi_i(k) N_Z(k) + 2Re \left[ Z_0^T(k) \Lambda N_Z(k) \right], \quad (G.4)
\]
\[
\frac{\partial V_2(\theta_0, \sigma^2_i, Z)}{\partial \sigma^2_{[i]}} |_{\sigma^2_0} = \sum_{k \in K} \left[ C_E^{-1}(k) \frac{\partial C_E(k)}{\partial \sigma^2_{[i]}} \right] + N_Z^T(k) W_i(k) N_Z(k). \quad (G.5)
\]
Utilizing (A.3), (G.4) and \( \operatorname{tr} [\Pi_i(k) \Sigma_0] = 0 \),
\[
(Q_1)_{[i,j]} = \sum_{k \in K} \operatorname{tr} [\Pi_i(k) \Sigma_0 \Pi_j(k) \Sigma_0] \\
+ \operatorname{tr} [\Pi_i(k) \Sigma_0 \Pi_j(k) \Sigma_0] + 2Re \left[ Z_0^T(k) \Lambda \Sigma_0 \Pi_j(k) Z_0(k) \right] \\
+ \sum_{k,l \in K} \operatorname{tr} [\Pi_i(k) \Sigma_0 \Pi_j(l) \Sigma_0] \\
= \left( M_1 \right)_{[i,j]} + \sum_{k \in K} \operatorname{tr} [\Pi_i(k) \Sigma_0 \Pi_j(k) \Sigma_0]. \quad (G.6)
\]
The expansion of the second term of (G.6) gives \( (M_2)_{[i,j]} \).

Considering (G.4), (G.5) and \( \operatorname{tr} [\Pi_i(k) \Sigma_0] = 0 \)
\[
(Q_2)_{[i,j]} = \sum_{k \in K} \operatorname{tr} [\Sigma_0 \Pi_i(k)] \sum_{l \in K} \operatorname{tr} \left[ C_E^{-1}(l, \vartheta_0) \frac{\partial C_E(l, \vartheta_0)}{\partial \sigma^2_{[i]}} \right] \\
+ \sum_{k \in K} \operatorname{tr} [\Pi_i(k) \Sigma_0 W_j(k) \Sigma_0] \\
+ \sum_{k,l \in K} \operatorname{tr} [\Pi_i(k) \Sigma_0 \Pi_j(l) \Sigma_0] \\
= \sum_{k \in K} \operatorname{tr} [\Pi_i(k) \Sigma_0 W_j(k) \Sigma_0], \quad (G.7)
\]
which equals zero by inserting (G.1) and (G.2).

Using (A.3) and (G.5),
\[
(Q_3)_{[i,j]} = \sum_{k \in K} \operatorname{tr} \left[ C_E^{-1}(k, \vartheta_0) \frac{\partial C_E(k, \vartheta_0)}{\partial \sigma^2_{[i]}} \sum_{l \in K} \operatorname{tr} [\Sigma_0 W_j(l)] \right] + \sum_{k,l \in K} \operatorname{tr} [\Sigma_0 W_i(k)] \sum_{l \in K} \operatorname{tr} \left[ C_E^{-1}(l, \vartheta_0) \frac{\partial C_E(l, \vartheta_0)}{\partial \sigma^2_{[j]}} \right] \\
+ \sum_{k \in K} \operatorname{tr} [W_i(k) \Sigma_0] \sum_{l \in K} \operatorname{tr} [W_j(l) \Sigma_0] \\
+ \sum_{k \in K} \operatorname{tr} [W_i(k) \Sigma_0] \sum_{l \in K} \operatorname{tr} [W_j(l) \Sigma_0]
\]
10
\[ + \sum_{k \in \mathcal{K}} \text{tr} \left[ W_i(k) \Sigma_0 W_j(k) \Sigma_0 \right], \quad (G.8) \]

which gives \((M_3)_{ij}\) by considering that

\[ \text{tr} \left[ W_i(k) \Sigma_0 \right] = -\text{tr} \left[ C_{E}^{-1}(k, \vartheta_0) \frac{\partial C_{E}(k, \vartheta_0)}{\partial \sigma^2_{i}} \right]. \quad (G.9) \]

**References**


