Asymptotic Analysis of Semi-Parametric Weighted Null-Space Fitting Identification

Miguel Galrinho, Cristian R. Rojas, Member, IEEE, and Håkan Hjalmarsson, Fellow, IEEE

Abstract—Standard system identification methods often provide biased estimates with closed-loop data. With the prediction error method (PEM), the bias issue is solved by using a noise model that is flexible enough to capture the noise spectrum. However, a too flexible noise model (i.e., too many parameters) can cause additional numerical problems for PEM. In this paper, we perform a theoretical analysis of the weighted null-space fitting (WNSF) method when a parametric noise model is not estimated. With this method, the system is first captured using a non-parametric ARX model, which is then reduced to a parametric model of interest. In the reduction step, a noise model does not need to be estimated if it is not of interest. In open loop, this still provides asymptotically efficient estimates of the dynamic model. In closed loop, the estimates are consistent, and their covariance is optimal for a non-parametric noise model. In this paper, we prove these results, which require additional technical details compared with the case with a full parametric model structure. In particular, we use a geometric approach for variance analysis, deriving a new result that will be instrumental to our end. Finally, we use a simulation study to illustrate the benefits of the method when the noise model cannot be parametrized by a low-order model.

Index Terms—System identification, least squares

I. INTRODUCTION

The prediction error method (PEM) is a benchmark for estimation of linear parametric models. If the model orders are correct and the noise is Gaussian, PEM with a quadratic cost function is asymptotically efficient: the asymptotic covariance of the estimates is the Cramér-Rao (CR) bound—the lowest covariance attainable by a consistent estimator.

Two models can typically be distinguished in a parametric model structure: the dynamic model and the noise model. Because the noise sequence is often the result of different noise contributions aggregated in a complex or even intractable manner, the concept of a “correct order” for the noise model is often inappropriate in practice. If the noise-model order is chosen too small with PEM, the dynamic-model estimate will be consistent in open loop, but biased in closed loop.

The bias issue with closed-loop data is not exclusive of PEM. Instrumental variable methods [2] require the reference signal to construct the instruments in closed loop [3,4]. For classical subspace methods [5,6], the bias issue in closed loop has been overcome by more recent algorithms [7,8].

With PEM, the bias issue can in theory be solved by letting the noise model structure be more flexible (i.e., letting the number of parameters become very large), guaranteeing that the noise spectrum is captured. If the global minimum of the PEM cost function is found, in open loop this will asymptotically not affect the dynamic-model estimates; in closed loop, consistency is attained but not efficiency. The problem is that, because the noise model might require many parameters, the PEM cost function will potentially have more local minima, thus aggravating the numerical search for the global minimum.

Some methods use a semi-parametric approach, estimating a non-parametric noise model estimated in a first step, which is then used in a second step to estimate the dynamic model. Possible approaches have been delineated both in the frequency [9]–[11] and time domains [12]–[14].

The weighted null-space fitting (WNSF) method [15] also first estimates a non-parametric model, and then reduces it to a full parametric model (i.e., dynamic and noise models). Moreover, WNSF does not apply non-linear optimization techniques, but uses weighted least squares iteratively. In this sense, the method can be seen as belonging to the family of iterative least-squares methods. These methods date back to [16], and have later been denoted as iterative quadratic maximum likelihood (IQML) methods, with applications to filter design [17,18] and identification of dynamical systems [19]–[21]. The Steiglitz-McBride method [22] belongs also to this class of methods, being equivalent to IQML for an impulse-input case [23]. However, unlike in the aforementioned identification works, WNSF is asymptotically efficient in one iteration in open and closed loop for Box-Jenkins systems [24].

Instead of simultaneously estimating the dynamic and noise models, WNSF also allows to disregard the parametric noise model, reducing the non-parametric model estimate to obtain a parametric dynamic model only. This guarantees asymptotically efficient estimates in open loop, and consistent estimates in closed loop with optimal asymptotic covariance when an infinite-order noise model is used. This distinguishes this method from [13,14], which are developed for open loop. The asymptotic properties of the proposed method correspond to the asymptotic properties of PEM with an infinite-order noise model [25] in both open and closed loop, but performed with a robust numerical procedure.

In [24], an extensive study of the fully parametric WNSF method has been conducted, including theoretical analysis and simulations. The case where WNSF is used with no parametric noise-model estimate—denoted semi-parametric WNSF—has been addressed and illustrated in [24], but a formal proof of the asymptotic properties has not been provided. In this paper, we perform the analysis of this setting, which has

Automatic Control Lab and ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: [galrinho, crrro, hjalmars]@kth.se.)
This work was supported by the Swedish Research Council under contracts 2015-05285 and 2016-06079.
additional difficulties because of the non-squared dimension of some matrices. For this purpose, we use the geometric approach of [26], extending the results therein to the setting of our problem. Thus, our contributions are: first, we derive a result for variance analysis in system identification; second, we use this result to perform a theoretical analysis of the semi-parametric WNSF method, proving consistency and deriving the asymptotic covariance of the estimates in open and closed loop. This analysis is complemented with a simulation study to illustrate the benefits of the method.

II. PRELIMINARIES

Essentially, the same notation, definitions, and assumptions used in [23] apply to this paper. For convenience, we reproduce the notation and assumptions here; for the definitions of $f_N$-stability and $f_N$-quasi-stationarity, see [24][27].

A. Notation

- $\|x\| = \sqrt{\sum_{k=1}^{n} |x_k|^2}$, with $x$ an $n \times 1$ vector.
- $\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\|$, with $A$ a matrix and $x$ a vector of appropriate dimensions.
- $\|A\|_F = \sqrt{\text{Trace}(AA^*)}$ (i.e., the Frobenius norm), with $A$ a matrix.
- $\|G(q)\|_{\mathcal{H}_\infty} := \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Trace}G(e^{i\omega})G^*(e^{i\omega})d\omega}$, with $G(q)$ a transfer matrix.
- $\|G(q)\|_{\mathcal{H}_\infty} := \sup_{\omega} \|G(e^{i\omega})\|$.
- $C$ denotes any constant, which need not be the same in different expressions.
- $\Gamma_n(q) = \begin{bmatrix} q^{-1} & \cdots & q^{-n} \end{bmatrix}^T$, where $q^{-1}$ is the backward time-shift operator.
- $A^*$ is the complex conjugate transpose of the matrix $A$.
- $\mathcal{T}_{n,m}(X(q))$ is the lower-triangular Toeplitz matrix of size $n \times m$ ($m \leq n$) with first column $[x_0 \cdots x_{n-1}]^T$, where $X(q) = \sum_{k=0}^{\infty} x_k q^k$. The dimension $n$ may be infinity, denoted $\mathcal{T}_{\infty,m}(X(q))$.

$E[x]$ denotes expectation of the random vector $x$.

$\mathbb{E}_{x} : = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}_x x_k$.

$x_N = \hat{O}(N)$ means that the function $x_N$ tends to zero at a rate not slower than $N$, as $N \to \infty$, w.p.1.

$\langle X(q), Y(q) \rangle := \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{i\omega})Y^*(e^{i\omega})d\omega$, where $X(q)$ and $Y(q)$ are transfer matrices of appropriate sizes.

B. Assumptions

Assumption 1 (True system and parametric model). The system has scalar input $\{u_t\}$, scalar output $\{y_t\}$, and is subject to scalar noise $\{e_t\}$. These signals are related by

$$y_t = G_o(q)u_t + H_o(q)e_t,$$

where $G_o(q)$ and $H_o(q)$ are rational functions according to

$$G_o(q) = \frac{L_o(q)}{F_o(q)} = \frac{\sum_{l=0}^{r} l^q q^{-l} + \cdots + l^m q^{-m}}{1 + f_1 q^{-1} + \cdots + f_m q^{-m}},$$

$$H_o(q) = \frac{C_o(q)}{D_o(q)} = \frac{1 + c^q q^{-1} + \cdots + c^m q^{-m}}{1 + d^q q^{-1} + \cdots + d^m q^{-m}}.$$

The transfer functions $G_o$, $H_o$, and $H_o^{-1}$ are assumed to be stable. The polynomials $L_o$ and $F_o$—as well as $C_o$ and $D_o$—do not share common factors.

We parametrize $G_o(q)$ as

$$G(q, \theta) = \frac{L(q, \theta)}{F(q, \theta)} = \frac{l_1 q^{-1} + \cdots + l_m q^{-m}}{1 + f_1 q^{-1} + \cdots + f_m q^{-m}},$$

where

$$\theta = [f_1 \cdots f_m l_1 \cdots l_m]^T$$

is the parameter vector to estimate, with known orders $m_f$ and $m_l$. The noise model orders $m_c$ and $m_d$ are not known.

Assumption 2 (Noise). The noise sequence $\{e_t\}$ is a stochastic process that satisfies

$$E[e_t]\mathbb{E}_{\mathcal{F}_{t-1}} = 0, \quad E[e_t^2]\mathbb{E}_{\mathcal{F}_{t-1}} = \sigma^2, \quad E[\|e_t\|^2] \leq C, \forall t.$$

Assumption 3 (Input). The input sequence $\{u_t\}$ is defined by

$$u_t = -K(q)y_t + r_t$$

under the following conditions.

1) The sequence $\{r_t\}$ is assumed to be independent of $\{e_t\}$, $f_N$-quasi-stationary with $f_N = \sqrt{\log N}$, and uniformly bounded: $\|r_t\| \leq C, \forall t$.

2) Let the spectral factorization of the power spectral density of $\{r_t\}$ be $\Phi_r(z) = F_r(z)F_r(z^{-1})$. Then, $F_r(q)$ is assumed to be $f_N$-stable, with $f_N = \sqrt{N}$.

3) The closed loop system is $f_N$-stable with $f_N = \sqrt{N}$.

4) The feedback transfer function $K(z)$ is bounded on the unit circle.

5) The spectral density of the process $\{r_t e_t^T\}$ is bounded from below by the matrix $\delta I$, $\delta > 0$ (this implies an informative experiment).

Operation in open loop is obtained by taking $K(q) = 0$. Alternatively to (1), the true system can be written as

$$A_o(q)y_t = B_o(q)u_t + e_t,$$

where

$$A_o(q) := \frac{1}{H_o(q)} = 1 + \sum_{k=1}^{\infty} a_k q^k,$$

$$B_o(q) := \frac{G_o(q)}{H_o(q)} = \sum_{k=1}^{\infty} b_k q^k.$$

(3)

are stable (Assumption I). In an intermediate step, WNSF estimates truncated $A_o(q)$ and $B_o(q)$, using the ARX model

$$A(q, \eta^n)y_t = B(q, \eta^n)u_t + e_t,$$

(4)

where

$$\eta^n = [a_1 \cdots a_n b_1 \cdots b_n]^T,$$

(5)

$$A(q, \eta^n) = 1 + \sum_{k=1}^{n} a_k q^{-k}, \quad B(q, \eta^n) = \sum_{k=1}^{n} b_k q^{-k}.$$
D1. $n(N) \to \infty$, as $N \to \infty$;
D2. $n^{1+\delta}(N)/N \to 0$, for some $\delta > 0$, as $N \to \infty$.

C. Prediction Error Method

The prediction error method minimizes a cost function of the prediction errors

$$
\varepsilon_t(\theta, \varsigma) = H^{-1}(q, \varsigma) \left( y_t - \frac{L(q, \theta)}{F(q, \theta)} u_t \right),
$$

where $H(q, \varsigma)$ is some noise model parametrization, function of a parameter vector $\varsigma$. Using a quadratic cost function, which is optimal when the noise sequence is Gaussian, the PEM estimate of the parameters is obtained by minimizing

$$
J(\theta, \varsigma) = \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} \varepsilon_t^2(\theta, \varsigma),
$$

where $N$ is the sample size.

Let $H(q, \varsigma)$ be such that there exists $\varsigma = \varsigma_0$ for which $H(q, \varsigma_0) = H_{\theta}(q)$. Denoting by $\theta_{PEM}$ the parameter vector $\theta$ that (together with some $\varsigma$) minimizes (7), the estimate $\theta_{PEM}$ is asymptotically distributed as

$$
\sqrt{N}(\theta_{PEM} - \theta_0) \sim \mathcal{N}(0, \sigma^2_{\theta} M_{PEM}^{-1}),
$$

with (arguments often omitted for notational simplicity)

$$
\hat{\Omega} = \left[ \begin{array}{cc}
\hat{\Pi}_r & \hat{\Gamma}_m^f \\
\hat{\Pi}_f & \hat{\Gamma}_m
\end{array} \right]
$$

and $\Phi_r^+$ the spectrum of

$$
u^+_r = S_o(q)r_t,
$$

where $S_o(q) = [1 + K(q)G_o(q)]^{-1}$ is the sensitivity function.

In open loop (in which case $\Phi_r^+$ is simply the input spectrum), $M_{PEM} = M$, and it corresponds to the CR bound. In closed loop, $M_{PEM} = M$ when the number of parameters in $\varsigma$ tends to infinity. In this case, $M$ does not correspond to the CR bound, but to the best possible covariance (from a prediction error perspective) with a non-parametric noise model [25].

The interest of estimating a non-parametric noise model in closed loop is that even if the noise spectrum needs to be captured by a higher-order model, it will still be possible to obtain an estimate of the dynamic model $G(q, \theta)$. However, estimating a non-parametric noise model simultaneously with a parametric dynamic model with PEM is not realistic. The reason is that, as the number of parameters in $H(q, \varsigma)$ increases, the prediction error (6) becomes a more complicated function of $\varsigma$, and finding the global minimum of the non-convex cost function (7) is a more difficult task. Consequently, the result that PEM with a non-parametric noise model provides estimates with covariance corresponding to $M$ may not always be useful in practice. With WNSF, this setting can be handled without increasing the difficulty of the problem.

III. SEMI-PARAMETRIC WEIGHTED NULL-SPACE FITTING

The WNSF method consists of three steps [23]. First, we estimate a non-parametric ARX model, with least squares. Second, we reduce this estimate to a parametric model, with least squares. Third, we re-estimate the parametric model, with weighted least squares. We now detail the procedure for each step, without estimating a parametric noise model.

For the first step, consider (4) in the regression form

$$
y_t = (\varphi^n \varphi^n)^T \eta^n + \varepsilon_t,
$$

where

$$
\varphi^n = \begin{bmatrix}
-\eta_{t-1} & \cdots & -\eta_{-n} & u_{t-1} & \cdots & u_{t-n}
\end{bmatrix}^T.
$$

Then, the least-squares estimate of $\eta^n$ is obtained by

$$
\hat{\eta}_N^n = [R_N^n]^{-1}r_N^n,
$$

where

$$
R_N^n = \frac{1}{N} \sum_{t=n+1}^{N} \varphi_t^n(\varphi_t^n)^T, \quad r_N^n = \frac{1}{N} \sum_{t=n+1}^{N} \varphi_t^n y_t.
$$

The asymptotic distribution of the estimates is [27]

$$
\sqrt{N}(\hat{\eta}_N^n - \eta^n) \sim \mathcal{N}(0, \sigma^2_{\eta}[R^n]^{-1}),
$$

where

$$
R^n_N \to \bar{R} := \mathbb{E}[(\varphi^n \varphi^n)^T], \quad \text{as} \ N \to \infty, \ \text{w.p.1},
$$

$$
r^n_N \to \bar{r} := \mathbb{E}[\varphi^n y_t], \quad \text{as} \ N \to \infty, \ \text{w.p.1},
$$

$$
\hat{\eta}_N^n \to \eta^n := [\bar{R}]^{-1}\bar{r}, \quad \text{as} \ N \to \infty, \ \text{w.p.1}.
$$

For the second step, we obtain an estimate of $G(q, \theta)$, from the non-parametric ARX-model estimate. For this purpose, we use (2) and (3) to write

$$
F_o(q)B_o(q) - L_o(q)A_o(q) = 0.
$$

Because we are not interested in estimating a parametric noise model, we do not consider the part of (3) from where such a model could be obtained. By convolution, (12) can be written in matrix form as

$$
b^n_o - Q_n(\eta^n_o)\theta_0 = 0,
$$

where $\eta^n_o$ is given by (5) evaluated at the true coefficients of (3). $b^n_o$ consists of the last $n$ coefficients of $\eta^n_o$, and

$$
Q_n(\eta^n) = \left[ -Q^t_n(\eta^n) \ Q^t_n(\eta^n) \right],
Q^t_n(\eta^n) = T_{n,m}(A(q, \eta^n)), \quad Q^t_n(\eta^n) = T_{n,m}(B(q, \eta^n)).
$$

Motivated by (13), we replace $\eta^n_o$ by its estimate $\hat{\eta}_N^n$ (and the same for $b^n_o$, which is a part of $\eta^n_o$) and obtain an estimate of $\theta$ with least squares:

$$
\hat{\theta}_{LS}^N = (Q_N^n(\hat{\eta}_N^n)Q_N^n(\hat{\eta}_N^n))^{-1}Q_N^n(\hat{\eta}_N^n)b_N^n.
$$

For the third step, we re-estimate $\theta$ taking into account the errors in $\hat{\eta}_N^n$. As $\eta^n_o$ is replaced by $\hat{\eta}_N^n$ in (13), the residuals are given by $\hat{b}_N^n - Q_n(\hat{\eta}_N^n)\theta_0 = T_n(\theta_0)(\hat{\eta}_N^n - \eta^n_o)$, where

$$
T_n(\theta) = \begin{bmatrix}
-T^n_0(\theta) & T^n_1(\theta)
\end{bmatrix}, \quad T^n_0(\theta) = T_{n,n}(L(q, \theta)), \quad T^n_1(\theta) = T_{n,n}(F(q, \theta)).
$$
Theorem 1. Let Assumptions 1, 2, 3, and 4 hold, and Algorithm 1.

Proof. We will need some auxiliary results. For that, we will begin by writing, for the estimate from Step 3,

$$
\hat{\theta}_N^{\text{WLS}} - \theta_o = M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{\text{WLS}}) \{ Q_n^T(\hat{\eta}_N) W_n(\hat{\theta}_N^{\text{WLS}}) T_n(\theta_o) (\hat{\eta}_N - \eta_o^{(N)}) \},
$$

where $M(\eta^n, \theta) := Q_n^T(\eta^n) W_n(\theta) Q_n(\eta^n)$ and $\eta_N := \eta_o^{(N)}$, recalling that $n$ is a function of $N$ according to Assumption 4 (for notational simplicity, we use only $n$ instead of $n(N)$ in matrix subscripts). To study consistency and asymptotic distribution, the limit value of (19) and the asymptotic distribution of $\sqrt{N}(\hat{\theta}_N^{\text{WLS}} - \theta_o)$ will be analyzed. The challenge in this analysis (compared to (24)) is in how to treat the matrix $W_n(\hat{\theta}_N^{\text{WLS}})$, and consequently also the matrix $M(\eta_N, \hat{\eta}_N^{\text{WLS}})$, which contains $W_n(\hat{\theta}_N^{\text{WLS}})$. From (17), $W_n(\hat{\theta}_N^{\text{WLS}}) = [T_n(\hat{\theta}_N^{\text{WLS}}) R_n^T(\hat{\theta}_N^{\text{WLS}})]^{-1}$, which in (24) may be written as $W_n(\hat{\theta}_N^{\text{WLS}}) = T_n(\hat{\theta}_N^{\text{WLS}}) R_n^T(\hat{\theta}_N^{\text{WLS}})$, as the matrix $T_n(\hat{\theta}_N^{\text{WLS}})$ is square therein. However, in the semi-parametric version treated in this paper, $T_n(\hat{\theta}_N^{\text{WLS}})$ is not square, and we cannot analyze $W_n(\hat{\theta}_N^{\text{WLS}})$ by taking inverses of the individual matrices it consists of.

To deal with this issue, we use the approach in (26), writing the aforementioned matrices as projections of the rows of some matrix onto the subspace spanned by the rows of another matrix. This will be applied to the limit value of the matrix $M(\eta_N, \hat{\theta}_N^{\text{WLS}})$, defined by

$$
\bar{M}(\eta_o, \theta_o) := \lim_{n \to \infty} Q_n^T(\eta_o^n) [T_n(\theta_o) R_n^T(\theta_o)]^{-1} Q_n(\eta_o^n).
$$

Writing $Q_n(\eta_o^n)$, $T_n(\theta_o)$, and $R_n$ (defined in (14), (16), and (11), respectively) in the frequency domain as

$$
Q_n(\eta_o^n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \Gamma_m^0 \right] \Gamma_m^* \Gamma_m^* d\omega,
$$

$$
T_n(\theta_o) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \Gamma_n \right] \Gamma_n^* \Gamma_n^* d\omega,
$$

$$
R_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ \Gamma_n \right] \Gamma_n^* \Gamma_n^* d\omega,
$$

we may express $\bar{M}(\eta_o, \theta_o)$ as

$$
\bar{M}(\eta_o, \theta_o) = \lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \left[ \langle \Omega_n, \Omega_n \rangle \right]^{-1} \langle \Omega_n, \Psi_n \rangle \left[ \langle \Omega_n, \Psi_n \rangle \right]^{-1} \langle \gamma, \Psi_n \rangle.
$$

where

$$
\Omega_n = \left[ \begin{array}{c} -G_o(q) S_o(q) \Gamma_n(q) -\sigma_o H_o(q) \Gamma_n(q) \\ S_o(q) F_r(q) \Gamma_n(q) \end{array} \right],
$$

$$
\Psi_n = \left[ \begin{array}{c} F_o(q) \\ S_o(q) F_r(q) \end{array} \right],
$$

$$
\gamma = \left[ \begin{array}{c} -B_o(q) \Gamma_m(q) \\ A_o(q) \Gamma_m(q) \end{array} \right].
$$

Following the approach in (26), we recognize that the term $\langle \Psi_n, \Omega_n \rangle \left[ \langle \Omega_n, \Omega_n \rangle \right]^{-1} \langle \Omega_n, \Psi_n \rangle$ in (21) can be written as

$$
\langle \Omega_n, \Omega_n \rangle \left[ \langle \Omega_n, \Omega_n \rangle \right]^{-1} \langle \Omega_n, \Psi_n \rangle = \langle \text{Proj}_{\Omega_n} \Psi_n, \text{Proj}_{\Omega_n} \Psi_n \rangle,
$$

where $\text{Proj}_{\Omega_n} \Psi_n$ denotes the projection of the rows of $\Psi_n$ onto the subspace spanned by the rows of $\Omega_n$. As $n \to \infty$, the dimensions of the matrix $\Omega_n$ increase, and the subspace spanned by its rows approaches $H_2$. Then, the limit value of the projection will be the causal part of the projected matrix.
For a simplified case, suppose that $\Psi_n$ were causal and that its dimension did not depend on $n$ (i.e., $\Psi_n = \Psi$). In this case, we would have $\lim_{n \to \infty} \{\text{Proj}_{S_n}\Psi, \text{Proj}_{S_n}\Psi\} = \{\Psi, \Psi\}$. In turn, we would then have that $\hat{M}(\eta_0, \theta_0) = \langle \gamma, \Psi \rangle \langle \Psi, \gamma \rangle = \langle \text{Proj}_{S_n}\gamma, \text{Proj}_{S_n}\gamma \rangle$. If we now reintroduce that the dimension of $\Psi$ depends on $n$ ($\Psi_n = \Psi$), and we assume that the rows of $\Psi_n$ span $H_2$ as $n \to \infty$, we have that $\hat{M}(\eta_0, \theta_0) = \lim_{n \to \infty} \langle \text{Proj}_{S_n}\gamma, \text{Proj}_{S_n}\gamma \rangle = \langle \gamma, \gamma \rangle$. These arguments follow from results in [26]. However, handling the dimensional increase of $\Psi_n$ with $n$ requires additional technical developments. One of the key results for the asymptotic analysis in this paper is that the aforementioned result (i.e., that $\hat{M}(\eta_0, \theta_0) = \langle \gamma, \gamma \rangle$) still holds when the dimensions of $\Psi_n$ increase with $n$. This is considered in the following theorem.

**Theorem 2.** Let

$$
\Omega_n = \begin{bmatrix} F_1(q)\Gamma_n(q) & F_2(q)\Gamma_n(q) \\ F_3(q)\Gamma_n(q) & 0 \end{bmatrix},
\Psi_n = \begin{bmatrix} F_4(q)\Gamma_n(q) \\ 0 \end{bmatrix},
\gamma = \begin{bmatrix} F_5(q)\Gamma_{mf}(q) & 0 \\ F_6(q)\Gamma_{mf}(q) \end{bmatrix},
$$

where $F_j(q) = \sum_{k=0}^{\infty} f_k^{(j)} q^{-k}$ ($j = \{1, \ldots, 6\}$) are exponentially stable (i.e., $|f_k^{(j)}| < C\lambda^k \forall j, \lambda < 1$, $f_0^{(4)} \neq 0$, and $F_4^{-1}(q)$ and $F_5^{-1}(q)$ are exponentially stable. Then, if $||\Omega_n, \gamma|| < C \forall n$,

$$
\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \gamma \rangle^{-1} = \langle \gamma, \gamma \rangle.
$$

**Proof.** See Appendix B. \qed

**B. Consistency and Asymptotic Efficiency**

Using the auxiliary results derived above, we now show that semi-parametric WNSF is consistent and asymptotically efficient. Regarding consistency of Step 3 in Algorithm 1 we have the following result.

**Theorem 3.** Let Assumptions 1, 2, 3 and 4 hold. Then,

$$
\hat{\theta}_n^{WLS} \to \theta_0, \text{ as } N \to \infty, \text{ w.p.1.}
$$

**Proof.** See Appendix C. \qed

Theorem 3 implies that semi-parametric WNSF provides consistent estimates of $\theta_0$.

Regarding the asymptotic distribution and covariance of Step 3 in Algorithm 1 we have the following result.

**Theorem 4.** Let Assumptions 1, 2, 3 and 4 hold. Then,

$$
\sqrt{N}(\hat{\theta}_n^{WLS} - \theta_0) \sim \mathcal{N}(0, \sigma_{\theta}^2 M^{-1}),
$$

where $M$ is given by (8).

**Proof.** See Appendix D. \qed

As consequence of Theorem 4, the semi-parametric WNSF method summarized in Algorithm 1 is asymptotically efficient in open loop, with $M$ corresponding to the CR bound. In closed loop, $M$ corresponds to the best covariance attainable by PEM with an infinite-order noise model [23].

**V. SIMULATIONS**

In this section, we perform two simulation studies. In the first, we illustrate the asymptotic properties of the method. In the second, we illustrate how estimating a non-parametric noise model with WNSF may be useful in scenarios where a low-order parametrization for the noise model does not capture the noise spectrum accurately enough.

**A. Illustration of asymptotic properties**

According to the results in Section 4, semi-parametric WNSF is asymptotically efficient in open loop, with the asymptotic covariance of the dynamic-model estimates given by $\sigma_{\theta}^2 M^{-1}$. In closed loop, the asymptotic covariance is still given by $\sigma_{\theta}^2 M^{-1}$, but in this case it does not correspond to the CR bound, but to the optimal asymptotic covariance when the noise-model order tends to infinity. To illustrate this, we perform open- and closed-loop simulations such that the closed-loop data are generated by

$$
u_t = \frac{1}{1 + K(q)G_0(q)}r_t - \frac{1}{1 + K(q)G_0(q)}e_t,
\eta_t = \frac{G_0(q)}{1 + K(q)G_0(q)}r_t + \frac{H_0(q)}{1 + K(q)G_0(q)}e_t,
$$

and the open-loop data by

$$\nu_t = \frac{1}{1 + K(q)G_0(q)}r_t,
\eta_t = G_0(q)\nu_t + H_0(q)e_t,$$

where $\{r_t\}$ and $\{e_t\}$ are independent Gaussian white sequences with unit variance, $K(q) = 1$, and

$$G_0(q) = q^{-1} + 0.1q^{-2}, \quad H_0(q) = \frac{1 + 0.7q^{-1}}{1 - 0.4q^{-1}}.$$

We perform 1000 Monte Carlo runs, with sample sizes $N \in \{300, 600, 1000, 3000, 6000, 10000\}$. We apply WNSF with an ARX model of order 50 with the open- and closed-loop data. Performance is evaluated by the mean-squared error of the estimated parameter vector of the dynamic model,

$$\text{MSE} = ||\hat{\theta}_n^{WLS} - \bar{\theta}_0||^2,$$

As this simulation has the purpose of illustrating asymptotic properties, initial conditions are assumed known—that is, the spectrum of (9), is the same for both data sets, both scenarios have the same asymptotic covariance, in accordance to our theoretical results.

**B. Random noise model**

When a low-order parametrization of the noise model is not enough to capture the noise spectrum, the noise model may require many parameters. With PEM, a simultaneous
estimate of the dynamic model and a long noise model is not numerically robust due to the non-convexity of the cost function. The semi-parametric WNSF is appropriate to deal with this scenario, because the noise spectrum is captured beforehand with a non-parametric ARX model.

Modeling the correct noise spectrum is particularly important in closed loop, where the estimates of the plant will be biased if the noise model is not flexible enough to capture the noise spectrum. For this reason, we consider a closed-loop setting, where data are generated by

\[
\begin{align*}
u_t &= \frac{1}{1 + K(q)G_o(q)} r_t - \frac{K(q)H_o(q)}{1 + K(q)G_o(q)} e_t, \\
y_t &= \frac{G_o(q)}{1 + K(q)G_o(q)} r_t + \frac{H_o(q)}{1 + K(q)G_o(q)} e_t.
\end{align*}
\]

The signals \(\{r_t\}\) and \(\{e_t\}\) are Gaussian white noise sequences with variances 1 and 4, respectively. The system is given by

\[
G_o(q) = \frac{1.0q^{-1} - 0.80q^{-2}}{1 - 0.95q^{-1} + 0.90q^{-2}},
\]

the controller by \(K(q) = 0.2\), and the true noise model by

\[
H_o(q) = 1 + \sum_{k=1}^{n=10} \lambda_k q^{-k}
\]

with \(\lambda_k = w_k e^{-0.2k}\), where \(w_k\) is drawn from a Gaussian distribution with zero mean and unit variance. Here, differently than Assumption1 stability of \(H_o(q)\) is not ensured. However, this is not an issue if the noise is Gaussian, as there must exists an inversely stable \(H_o(q)\) for which the noise sequence has the same spectrum.

It may be difficult to find an appropriate low-order parametrization for the noise model (24) in the form

\[
H(q, \zeta) = \frac{c_1 q^{-1} + \cdots + c_{m_h} q^{-m_h}}{d_1 q^{-1} + \cdots + d_{m_h} q^{-m_h}},
\]

where

\[
\zeta = [c_1 \cdots c_{m_h} \ d_1 \cdots \ d_{m_h}]^T.
\]

In this case, a Box-Jenkins model is estimated. The most appropriate noise model order may be chosen by using some information criterion, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) [1].

The alternative is to use a high-order model for the noise model. For example,

\[
H(q, \zeta^n) = 1 + \sum_{k=1}^{n} \zeta_k q^{-k},
\]

or

\[
H(q, \zeta^n) = \frac{1}{1 + \sum_{k=1}^{n} \zeta_k q^{-k}},
\]

where

\[
\zeta^n = [\zeta_1 \cdots \zeta_n]^T.
\]

In particular, the choice (27) has the same structure as the noise model estimated by semi-parametric WNSF.

Motivated by these alternatives, we compare the following methods:

- semi-parametric WNSF, with non-parametric model order \(n = 200\);
- PEM, with default MATLAB initialization, and noise model (25) with \(m_h \in \{1, 2, \ldots, 30\}\), where the order \(m_h\) is chosen using the AIC or BIC criterion (denoted PEM\(_\text{aic}\) and PEM\(_\text{bic}\), respectively);
- PEM, with default MATLAB initialization, and noise model (27) with \(n = 200\) (denoted PEM\(_\text{np}\), where ‘np’ stands for non-parametric).

PEM uses the implementation in MATLAB2016b System Identification Toolbox. All the methods use a maximum of 100 iterations, but stop early upon convergence (default settings for PEM, 10\(^{-4}\) as tolerance for the normalized relative change in the parameter estimates for WNSF). The search algorithm used by PEM is chosen automatically. The noise model (26) was not used with PEM for computational reasons: the optimization becomes extremely slow as stability of the inverse of the noise model when computing the prediction errors (6) is difficult to fulfill with so many parameters, while the inverse of any estimate of (27) is always stable.

We use sample sizes \(N \in \{1000, 5000, 10000\}\) and perform 100 Monte Carlo runs. Performance is evaluated by the FIT of the impulse response of the dynamic model, given by

\[
\text{FIT} = 100 \left(1 - \frac{||g_o - \hat{g}||}{||g_o - \text{mean}(g_o)||}\right),
\]

in percent, where \(g_o\) is a vector with the impulse response parameters of \(G_o(q)\) (mean \(g_o\) is its average), and similarly for \(\hat{g}\) but for the estimated model. In (28), sufficiently long impulse responses are taken to make sure that the truncation of their tails does not affect the FIT.

The FITs for the different sample sizes are shown in Fig. 2. For \(N = 1000\), WNSF has the most robust performance, with smaller whiskers than the remaining algorithms and no low-performance occurrences, which occur when PEM is used with AIC/BIC or with a a non-parametric noise model. Among the PEM alternatives, an AIC/BIC criterion with a Box-Jenkins model with orders up to 30 performed better than using a non-parametric noise model. For \(N = 5000\), WNSF and PEM with non-parametric noise model have similar median performance, but the median performance was poorer than for WNSF and PEM\(_\text{np}\). Similar
conclusions can be drawn for \( N = 10000 \), where we observe that PEM does not necessarily provide better results with more data samples, potentially due to numerical problems.

Overall, WNSF showed more robust performance among the sample sizes used. However, an even more evident advantage is the computational time. Table I shows the average times, in seconds, required for the identification using WNSF, PEM with all the orders computed for AIC and BIC, and PEM with a parametric noise-model estimate, named the semi-parametric WNSF, for the different sample sizes (all the computations were performed in the same computer). Here, we observe how WNSF requires much lower computational time than the alternatives. This is a consequence of PEM estimating the noise model in the non-linear optimization procedure, while in WNSF the high-order model is estimated in a previous least-squares step, which is numerically robust. Moreover, the time required for WNSF and PEM with AIC/BIC does not change significantly with \( N \), unlike with PEM. In this case, the time does not necessarily decrease for smaller \( N \). The problem arising when using smaller \( N \) is that the cost function more likely becomes ill-conditioned at some parameter values during the optimization.

VI. CONCLUSION

Many standard system identification methods provide biased estimates with closed-loop data. In the particular case of PEM, the bias issue is avoided by choosing a noise-model order that is large enough to capture the noise spectrum. An appropriate order is often difficult to choose, and making it arbitrarily large increases the numerical problems of PEM. In [24], where WNSF is analyzed, it is stated that the method can be used without a parametric noise-model estimate, named the semi-parametric WNSF. In this paper, we deepen this discussion.

A simulation study illustrates the importance of separating the dynamic- and noise-model identification when a high-order noise model is required, both in terms of performance and computational time. With WNSF, this separation always occurs, as the method first estimates a high-order ARX model. Then, a low-order noise model does not need to be obtained, as the noise spectrum has been captured in the first step.

We also provide a theoretical analysis of the asymptotic properties. To this end, we extend the geometric approach in [26], deriving a more general result that may also be useful for variance analysis of other methods. We show that semi-parametric WNSF provides consistent estimates of the dynamic model with closed-loop data. With open-loop data, the estimates are also asymptotically efficient; with closed-loop data, the asymptotic covariance of the estimates corresponds to the best possible covariance with a non-parametric noise model. This gives WNSF attractive features in terms of flexibility of noise-model structure and asymptotic properties: if a correct parametric noise model is estimated, the dynamic-model estimates are asymptotically efficient; if not, they are consistent and optimal for a non-parametric noise model. We used a simulation study to illustrate how semi-parametric WNSF is an appropriate method for scenarios where the noise model cannot be accurately modeled with a low-order parametrization.

APPENDIX A

AUXILIARY RESULTS FOR THE PROOF OF THEOREM 2

The following results will be used to prove Theorem 2.

Extension of Cauchy-Schwarz inequality for transfer-matrix inner products: Let \( X(q) \) and \( Y(q) \) be transfer matrices and \( x \) and \( y \) be vectors of appropriate dimensions. Then, we have

\[
||\langle X, Y \rangle||^2 = \sup_{||x||=1, ||y||=1} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} x^T X(e^{i\omega}) Y(e^{i\omega}) y d\omega \right| 
\]

\[
\leq \sup_{||x||=1, ||y||=1} \left( x^T (X, X)x \ y^T (Y, Y)y \right)
\]

\[
= ||X, X|| \ ||Y, Y|| 
\]

(29)

Bound for spectral norm of inner product of transfer matrices: Let \( X(q) \) be a transfer matrix. Then, we have

\[
||\langle X, X \rangle|| \leq ||\langle X, X \rangle||_F = \sqrt{\text{Trace}(\langle X, X \rangle^2)} 
\]

\[
\leq \text{Trace}(X, X) = ||X||_H^2 \quad (30)
\]

where the second inequality follows from \( \text{Trace}(A^2) \leq [\text{Trace}(A)]^2 \) for a positive semi-definite matrix \( A \).

Bound for Toeplitz operators of stable filters: Let \( X(q) := \sum_{k=-\infty}^\infty x_k q^k \). From [28] Chapter 3), we have

\[
\left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_\infty \Gamma_\infty^* X(e^{i\omega}) d\omega \right| = ||X(q)||_{H_\infty} 
\]

\[
\geq \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_\eta \Gamma_\eta^* X(e^{i\omega}) \right| \quad (31)
\]

where the inequality is due to the matrix on the right-hand side of (31) being a block of the infinite-dimensional matrix on the left-hand side.

APPENDIX B

PROOF OF THEOREM 2

In this appendix, we prove Theorem 2.
Inner Projection: \( \langle \Psi, \Omega \rangle / (\Omega, \Omega)^{-1} (\Omega, \Psi) \)

Let
\[
\Psi = \Psi^c + \Psi^a,
\]
where
\[
\Psi^c := \begin{bmatrix}
 f_0^{(4)} e^{-i\omega} & 0 \\
f_0^{(4)} e^{-2i\omega} + f_1^{(4)} e^{-i\omega} & 0 \\
\vdots & \vdots \\
n_0^{(4)} e^{-n_k i\omega} + \sum_{k=1}^{n-1} f_n^{(4)} e^{-n_k i\omega} & 0
\end{bmatrix},
\]
\[
\Psi^a := \begin{bmatrix}
\sum_{k=0}^{n-1} f_k^{(4)} e^{i\omega k} & 0 \\
\sum_{k=0}^{n} f_k^{(4)} e^{i\omega k} & 0 \\
\vdots & \vdots \\
\sum_{k=0}^{n} f_k^{(4)} e^{i\omega k} & 0
\end{bmatrix}.
\]

Alternatively, \( \Psi^c_n \) can be written as \( \Psi^c_n = [P_n \Gamma_n] \), where \( P_n = T_n \times (F_4(q)) \). Using (32), we can write
\[
\langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) = \langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) + \langle \Psi^a_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^a_n) + \langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) + \langle \Psi^a_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^a_n) = \Lambda_n + \Delta_n^{(1)},
\]
where
\[
\Lambda_n = \langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) + \langle \Psi^a_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^a_n) + \langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) + \langle \Psi^a_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^a_n)
\]

By construction, \( \Psi^c_n(i) \) is a linear combination of the last \( n \) rows of \( \Omega_n \); therefore, \( \Psi^c_n(i) \in S_{\Omega_n} \).

Using (33), (34), and (35), we write
\[
\langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) = \langle \Psi^c_n, \hat{\Psi}^c_n \rangle + \Delta_n^{(1)}.
\]

Moreover, we can re-write (36) as
\[
\langle \Psi^c_n, \Omega_n \rangle / (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi^c_n) = \langle \Psi^c_n, \hat{\Psi}^c_n \rangle + \Delta_n^{(1)},
\]
where \( \Delta_n = \Delta_n^{(1)} + \Delta_n^{(2)} \) and
\[
\Delta_n^{(2)} = \langle \Psi^c_n - \hat{\Psi}^c_n, \Psi^c_n \rangle + \langle \Psi^c_n - \hat{\Psi}^c_n, \Psi^c_n - \hat{\Psi}^c_n \rangle + \langle \Psi^c_n - \hat{\Psi}^c_n, \Psi^c_n - \hat{\Psi}^c_n \rangle.
\]

Replacing (37) in (23), we obtain
\[
\lim_{n \to \infty} \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle)^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma) = \lim_{\gamma \to \infty} \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma),
\]
and using the Sherman-Morrison-Woodbury formula, we rewrite (38) as
\[
\lim_{n \to \infty} \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma) = \lim_{n \to \infty} \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma) + \lim_{n \to \infty} \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} \Delta_n
\]
\[
\cdot (I + \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma), \langle \gamma, \Psi^c_n \rangle)^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma).
\]

We want to show that the second term on the right-hand side of (39), for which we can write
\[
\lim_{n \to \infty} \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle)^{-1} \Delta_n[I + \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle)^{-1} \Delta_n]^{-1}
\]
\[
\cdot (I + \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma), \langle \gamma, \Psi^c_n \rangle)^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma)
\]
\[
\leq \|(\gamma, \Psi^c_n)\| \|\langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma)\| \|(I + \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma), \langle \gamma, \Psi^c_n \rangle)^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma)\|
\]
\[
\leq \|(\gamma, \Psi^c_n)\| \|\langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma)\| \|(I + \langle \gamma, \Psi^c_n \rangle / (\langle \gamma, \Psi^c_n \rangle + \Delta_n^{(1)})^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma), \langle \gamma, \Psi^c_n \rangle)^{-1} (\langle \gamma, \Psi^c_n \rangle, \gamma)\|
\]
tends to zero as \( n \) tends to infinity. We start by considering the term \( \Delta_n \), for which we will need that
\[
\langle \Psi^c_n, \Psi^c_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_n \Gamma_n F_n P_n^T d\omega = P_n P_n^T
\]

Using also (29) and (30), we can write
\[
\|\Delta_n\| \leq 2\|\langle \Omega_n, \Omega_n \rangle \| \|\Omega_n, \Omega_n \|^{-1} \|P_n\| \|\Psi^c_n - \hat{\Psi}^c_n\| \|\hat{\Psi}^c_n\|
\]
\[
+ 2\|\langle \Omega_n, \Omega_n \rangle \| \|\Omega_n, \Omega_n \|^{-1} \|\Psi^c_n - \hat{\Psi}^c_n\| \|\hat{\Psi}^c_n\|
\]
\[
+ 2\|\langle \Omega_n, \Omega_n \rangle \| \|\Omega_n, \Omega_n \|^{-1} \|\Psi^c_n - \hat{\Psi}^c_n\| \|\hat{\Psi}^c_n\|
\]
\[
\leq \|F_n\| \sum_{k=0}^{n} \beta_k q^k - \sum_{k=0}^{n} \beta_k q^k \leq C\lambda^n
\]

Then,
\[
\|\Psi^c_n - \hat{\Psi}^c_n\| \|\hat{\Psi}^c_n\| \leq C\sqrt{n} \lambda^n \to 0 \text{ as } n \to \infty.
\]

Moreover, using (29), we have
\[
\|P_n\| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_n d\omega \right| \leq \|F_n\| \|\Gamma_n\| = C,
\]

and (by assumption) \|\langle \Omega_n, \Omega_n \rangle\| \leq C. Together with (42) and (43),
\[
\|\Delta_n\| \to 0 \text{ as } n \to \infty.
\]

Then, if the remaining matrix norms in the right-hand side of (40) are bounded, this term will tend to zero as \( n \to \infty \). For \( \langle \gamma, \Psi^c_n \rangle \), we have
\[
\|\langle \gamma, \Psi^c_n \rangle\| \leq \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_n F_n d\omega \right| \leq \|F_n\| \|\Gamma_n\| = C.
\]
Also, the inverse of $\langle \Psi_n^c, \Psi_n^c \rangle$ exists because $P_n$ is lower triangular with non-zero diagonal entries, and
\[
||\langle \Psi_n^c, \Psi_n^c \rangle^{-1}|| = P_n^{-1} F_n^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_n^{-1} \, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_n \Gamma_n^* F_n^{-1} \, d\omega \leq \|F_n\|^2_{\infty} = C.
\]
Using (39, 40, 45, 46), and (47), we have
\[
\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \langle \Omega_n, \Omega_n \rangle^{-1} \langle \Omega_n, \Psi_n \rangle^{-1} = \lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = 0.
\]

**Outer projection:** $\langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = \langle \gamma, \Psi_n \rangle$:
Recalling that $\Psi_n$ can be written as (32), we use that $\gamma$ is causal and $\Psi_n^c$ anti-causal to conclude that, analogously to (33),
\[
\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = \langle \gamma, \Psi_n \rangle.
\]
Now, row $i$ of $\gamma$ can be written as $\gamma(i) =: \{ \sum_{k=1}^{\infty} s_{k}^{(i)} q^{-k} \}$, where $|s_{k}^{(i)}| \leq C \lambda^{-k}$, $\lambda < 1$ due to exponential stability. Let also $\gamma_n(i) := \{ \sum_{k=1}^{\infty} s_{k}^{(i)} q^{-k} \}$ be row $i$ of a matrix $\gamma_n$. We re-write the right side of (49) as
\[
\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = \lim_{n \to \infty} \langle \gamma_n, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} + \text{lim}_{n \to \infty} \Delta_n^{(3)},
\]
where
\[
\Delta_n^{(3)} = \langle \gamma - \gamma_n, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} + \langle \gamma_n, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} + \langle \gamma - \gamma_n, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = 0.
\]
Using a similar approach for $\Delta_n^{(3)}$ as we did for $\Delta_n$, it can be shown that
\[
\|\Delta_n^{(3)}\| \leq C (\|\gamma - \gamma_n\|_{H_2} + \|\gamma - \gamma_n\|^2_{H_2}) \to 0 \text{ as } n \to \infty.
\]
Thus, (50) reduces to
\[
\lim_{n \to \infty} \langle \gamma, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = \lim_{n \to \infty} \langle \gamma_n, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = 0.
\]
Finally, we have that
\[
\langle \gamma_n, \Psi_n \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = \langle \text{Proj}_{S_{\Psi_n}} \gamma_n, \text{Proj}_{S_{\Psi_n}} \gamma_n \rangle = \gamma_n, \gamma_n=n,
\]
where the last equality follows from $\gamma_n \in S_{\Psi_n}$, as consequence of $P_n$ being invertible. Thus, replacing (33) in (52), we have
\[
\lim_{n \to \infty} \langle \gamma, \Psi_n^c \rangle \langle \Psi_n^c, \Psi_n^c \rangle^{-1} = \lim_{n \to \infty} \langle \gamma_n, \gamma_n \rangle = \langle \gamma, \gamma \rangle,
\]
which, together with (48), (49), (50), and (51) implies (23), as we wanted to show.

**APPENDIX C**

**Proof of Theorem 2**

We will show that
\[
||\dot{\theta}_{N}^{\text{WLS}} - \theta_0|| \to 0, \text{ as } N \to \infty, \text{ w.p.1.}
\]
To do this, we use (19) to write
\[
||\dot{\theta}_{N}^{\text{WLS}} - \theta_0|| \leq ||M^{-1}(\hat{\eta}_N, \hat{\theta}_S^{N})|| \cdot ||Q_n(\hat{\eta}_N)|| \cdot \left|\left| W_n(\hat{\theta}_S^{N}) \right|\right| \cdot \left|\left| T_n(\theta_0) \right|\right| \cdot \left|\left| \eta - \eta_0^{\infty}(N) \right|\right|.
\]
From (24), we have that
\[
||\hat{\eta}_N - \eta_0^{\infty}(N)|| \to 0 \text{ as } N \to \infty, \text{ w.p.1,}
\]
and that, w.p.1, there exists $\bar{N}$ such that
\[
||Q_n(\hat{\eta}_N)|| \leq C \forall N > \bar{N}.
\]
Then, we have left to show that $W_n(\hat{\theta}_S^{N})$ is bounded and $M(\hat{\eta}_N, \hat{\theta}_S^{N})$ is invertible for sufficiently large $N$.
We begin by considering the inverse of $W_n(\hat{\theta}_S^{N})$, for which we have
\[
\left|\left| T_n(\hat{\theta}_S^{N}) R_n^{-1} T_n^T(\hat{\theta}_S^{N}) \right|\right| \leq \left|\left| T_n(\theta_0) \bar{R}_n^{-1} T_n^T(\theta_0) \right|\right| + \left|\left| T_n(\hat{\theta}_S^{N}) R_n^{-1} T_n^T(\hat{\theta}_S^{N}) - T_n(\theta_0) \bar{R}_n^{-1} T_n^T(\theta_0) \right|\right|.
\]
In turn, it can be shown that
\[
T_n(\theta_0) [\bar{R}_n^{-1} T_n^T(\theta_0)] = [\Psi_n, \Omega_n] (\Omega_n, \Omega_n)^{-1} (\Omega_n, \Psi_n),
\]
where $\Psi_n$ and $\Omega_n$ are given by (22), which satisfy the conditions of Theorem 2 with $F_1(q) = -C_0(q) S_0(q)$, $F_2(q) = -\sigma_0 H_0(q)$, $F_3(q) = S_0(q) F_0(q)$, and $F_4(q) = F_0(q)/[S_0(q) F_0(q)]$. From (27), (41), (42), and (45), we have that there is $\bar{n}$ such that
\[
\left|\left| T_n(\theta_0) \bar{R}_n^{-1} T_n^T(\theta_0) \right|\right| = \left|\left| \Psi_n^c + \Delta_n \right|\right| \leq C \forall n > \bar{n}.
\]
Then, and using also (47), we have
\[
||\bar{W}(\theta_0)|| := ||T_n(\theta_0) \bar{R}_n^{-1} T_n^T(\theta_0)||^{-1} = \left|\left| \Psi_n^c + \Delta_n \right|\right|^{-1} \leq C \forall n > \bar{n}.
\]
Concerning the second term on the right-hand side of (59), we can write
\[
\left|\left| T_n(\hat{\theta}_S^{N}) R_n^{-1} T_n^T(\hat{\theta}_S^{N}) - T_n(\theta_0) \bar{R}_n^{-1} T_n^T(\theta_0) \right|\right| \leq \left|\left| T_n(\hat{\theta}_S^{N}) - T_n(\theta_0) \right|\right| \left|\left| R_n^{-1} \right|\right| \left|\left| T_n(\hat{\theta}_S^{N}) \right|\right| + \left|\left| T_n(\hat{\theta}_S^{N}) - T_n(\theta_0) \right|\right| \left|\left| R_n^{-1} \right|\right| \left|\left| T_n(\theta_0) \right|\right| + \left|\left| T_n(\theta_0) \right|\right|^2 \left|\left| [R_n^{-1} - \bar{R}_n^{-1}] \right|\right|.
\]
Now, the results in (24) apply to (62). In particular, there is $\bar{N}$ such that
\[
||[R_n^{-1}]^{-1} \leq C \forall N > \bar{N}, \quad ||R_n^0|| \leq C \forall N > \bar{N},
\]
w.p.1, and that
\[
\left|\left| T_n(\hat{\theta}_S^{N}) - T_n(\theta_0) \right|\right| \to 0 \text{ as } N \to \infty, \text{ w.p.1.}
\]
Together with (57), we conclude that
\[ ||T_n(\hat{\theta}_N^{LS})[R_n^\top]^{-1}T_n^\top(\hat{\theta}_N^{LS}) - T_n(\theta_0)[\hat{R}_n^\top]^{-1}T_n^\top(\theta_0)|| \rightarrow 0 \] as \( N \to \infty \), w.p.1. (64)

Using (63), (60), and (59), there is \( \bar{N} > 0 \) such that
\[ ||T_n(\hat{\theta}_N^{LS})[R_n^\top]^{-1}T_n^\top(\hat{\theta}_N^{LS})|| \leq C \quad \forall N > \bar{N} \] w.p.1.

Because of (63) and invertibility of \( T_n(\theta_0)[\hat{R}_n^\top]^{-1}T_n^\top(\theta_0) \), by continuity of eigenvalues there is \( \bar{N} > 0 \) such that \( W(\hat{\theta}_N^{LS}) = [T_n(\hat{\theta}_N^{LS})[R_n^\top]^{-1}T_n^\top(\hat{\theta}_N^{LS})]^{-1} \) exists for all \( N > \bar{N} \), and
\[ ||W(\hat{\theta}_N^{LS}) - \bar{W}(\theta_0)|| \rightarrow 0, \text{ as } N \to \infty, \text{ w.p.1.} \] (65)

Moreover, (65) and (61) imply that, w.p.1,
\[ ||W(\hat{\theta}_N^{LS})|| \leq C \quad \forall N > \bar{N}. \] (66)

Having shown (66), we have left to show that \( M(\hat{\eta}_N, \hat{\theta}_N^{LS}) \) is invertible for sufficiently large \( N \), in order to show (54).

Recall the definition (20), which can alternatively be written as (21), where \( \gamma \) is given by (22). Then, from Theorem 2 with
\[ F_0(q) = -\frac{B_o(q)S_0(q)F_r(q)}{F_o(q)}, \quad F_0(q) = \frac{A_o(q)S_0(q)F_r(q)}{F_o(q)}, \]
we have that
\[ M(\eta_0, \theta_0) = M. \] (67)

where \( M \) is given by (3). Because the inverse of \( M \) corresponds to the CR bound for an open-loop problem with input \( u_t = S_o(q)r_t \), and the CR bound exists for an informative experiment \( \mu \), we conclude that \( \hat{M}(\eta_0, \theta_0) \) is invertible. Then, we analyze the difference
\[ ||M(\hat{\eta}_N, \hat{\theta}_N^{LS}) - Q_n^\top(\eta_0^n)W_n(\theta_0)Q_n(\eta_0^n)|| \leq ||Q_n(\hat{\eta}_N) - Q_n(\eta_0^n)|| \cdot ||W(\hat{\theta}_N^{LS})|| \cdot ||Q_n(\eta_0^n)|| \]
\[ + ||Q_n(\hat{\eta}_N) - Q_n(\eta_0^n)|| \cdot ||W(\hat{\theta}_N^{LS})|| \cdot ||Q_n(\eta_0^n)|| \]
\[ + ||Q_n(\eta_0^n)|| \cdot ||W(\hat{\theta}_N^{LS}) - \bar{W}(\theta_0)|| \]

From (24), we have that
\[ ||Q_n(\hat{\eta}_N) - Q_n(\eta_0^n)|| \rightarrow 0, \text{ as } N \to \infty, \text{ w.p.1.} \] (68)

Together with (69) and (55), we conclude that
\[ ||M(\hat{\eta}_N, \hat{\theta}_N^{LS}) - Q_n^\top(\eta_0^n)W_n(\theta_0)Q_n(\eta_0^n)|| \rightarrow 0, \text{ as } N \to \infty, \text{ w.p.1.} \] (69)

Using (69), invertibility of \( \hat{M}(\eta_0, \theta_0) \), and continuity of eigenvalues, we have that there is \( \bar{N} > 0 \) such that \( M(\hat{\eta}_N, \hat{\theta}_N^{LS}) \) is invertible for all \( N > \bar{N} \),
\[ ||M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{LS})|| \leq C \quad \forall N > \bar{N}, \] (70)

and, using also (67) and (20),
\[ M^{-1}(\hat{\eta}_N, \hat{\theta}_N^{LS}) \rightarrow M^{-1} \quad \text{as } N \to \infty, \text{ w.p.1.} \] (71)

Finally, using (70), (66), (58), (57), (56), and (55), we conclude that (54) is satisfied, as we wanted to show. □
and because (66) and (61) guarantee that \( \hat{W}_n(\theta_0) \) and \( W_n(\hat{\theta}_N^{SL}) \) are bounded (in the latter, for sufficiently large \( N \)), it follows from (76) that
\[
\|\sqrt{N}Q_n^T(\hat{\theta}_N^{SL})W_n(\theta_0)T_n(\theta_0)(\hat{\eta}_N - \hat{\eta}^{p(N)})\| \leq C\sqrt{N}\|\hat{W}_n^{-1}(\hat{\theta}_N^{SL}) - W_n^{-1}(\hat{\theta}_N^{SL})\|\|\hat{\eta}_N - \hat{\eta}^{p(N)}\|.
\]

Now, the term \( \|W_n^{-1}(\hat{\theta}_N^{SL}) - W_n^{-1}(\theta_0)\| \) corresponds to (62); so, using (24)
\[
\|T_n(\hat{\theta}_N^{SL}) - T_n(\theta_0)\| = O\left(\sqrt{n^2(N)\log \frac{N}{N}(1 + d(N))}\right),
\]
(77) and (63), the first two terms on the right-hand side of (62) decay with (77). For the third term, we first write
\[
\|R_n^{-1} - [\hat{R}_n^{-1}]\| \leq \|\hat{R}_n^{-1} - R_n^{-1}\| \|R_n^{-1}\|^{-1}.
\]
In (27), it is shown that
\[
\|R_n^{-1} - R_n\| = O\left(2\sqrt{n^2(N)\log \frac{N}{N}} + Cn^2(N)\right).
\]
Then, using also (78), (63), and (77), we have that the third term on the right-hand side of (62) decays according to (79).
Thus, we have that
\[
\|W_n(\hat{\theta}_N^{SL}) - W_n(\theta_0)\| = O\left(2\sqrt{n^2(N)\log \frac{N}{N}}\right),
\]
(80) and (27)
\[
\|\hat{\eta}_N - \hat{\eta}^{p(N)}\| = O\left(\sqrt{n(N)\log \frac{N}{N}(1 + d(N))}\right),
\]
considering only the slowest-decaying term. Then, from (76), (80), and (27)
\[
\sqrt{N}Q_n^T(\hat{\theta}_N^{SL})W_n(\theta_0)T_n(\theta_0)(\hat{\eta}_N - \hat{\eta}^{p(N)})
\]
(81) and (24)—and, in turn, \( x(\theta_0; \hat{\eta}_N, \hat{\theta}_N^{SL}) \)—have the same asymptotic distribution and covariance. Thus, we will analyze (81) instead of \( x(\theta_0; \hat{\eta}_N, \hat{\theta}_N^{SL}) \).

Applying (27) Theorem 7.3] to (81) with \( Y_n = Q_n^T(\hat{\theta}_N^{SL})W_n(\theta_0)T_n(\theta_0) \)—and recalling that it has the same asymptotic distribution and covariance as \( x(\theta_0; \hat{\eta}_N, \hat{\theta}_N^{SL}) \)—we have that \( x(\theta_0; \hat{\eta}_N, \hat{\theta}_N^{SL}) \) is distributed according to (72) with
\[
X = \lim_{n \to \infty} Q_n^T(\hat{\theta}_N^{SL})W_n(\theta_0)T_n(\theta_0)\sigma_\theta^2
\]
\[
\cdot \hat{R}_n^0T_n(\theta_0)\hat{W}_n(\theta_0)Q_n(\hat{\eta}_N^{p(N)}) = \sigma_\theta^2 \lim_{n \to \infty} Q_n^T(\hat{\theta}_N^{SL})W_n(\theta_0)Q_n(\hat{\eta}_N^{p(N)}) = \sigma_\theta^2 M,
\]
where the last equality follows from (20). Then, replacing (82) in (73), we obtain
\[
\sqrt{N}(\hat{\theta}_N^{SL} - \theta_0) \sim \text{As}N\left(0, M^{-1}\right).
\]
[27] L. Ljung and B. Wahlberg. Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. Advances in Applied Probabilities, 24:412–440, 1992.

[28] V. Peller. Hankel operators and their applications. Springer, 2003.

[29] T. Söderström and P. Stoica. System Identification. Prentice Hall, 1989.

Miguel Galrinho was born in 1988. He received his M.S. degree in aerospace engineering in 2013 from Delft University of Technology, The Netherlands, and the Licentiate degree in automatic control in 2016 from KTH Royal Institute of Technology, Stockholm, Sweden.

He is currently a PhD student at KTH, with the Department of Automatic Control, School of Electrical Engineering, under supervision of Professor Håkan Hjalmarsson. His research is on least-squares methods for identification of structured models.

Cristian R. Rojas (M’13) was born in 1980. He received the M.S. degree in electronics engineering from the Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2004, and the Ph.D. degree in electrical engineering at The University of Newcastle, NSW, Australia, in 2008.

Since October 2008, he has been with the Royal Institute of Technology, Stockholm, Sweden, where he is currently Associate Professor at the Department of Automatic Control, School of Electrical Engineering. His research interests lie in system identification and signal processing.

Håkan Hjalmarsson (M’98–SM’11–F’13) was born in 1962. He received the M.S. degree in electrical engineering in 1988, and the Licentiate and Ph.D. degrees in automatic control in 1990 and 1993, respectively, all from Linköping University, Linköping, Sweden.

He has held visiting research positions at California Institute of Technology, Louvain University, and at the University of Newcastle, Australia. His research interests include system identification, signal processing, control and estimation in communication networks and automated tuning of controllers.

Dr. Hjalmarsson has served as an Associate Editor for Automatica (1996–2001) and for the IEEE Transactions on Automatic Control (2005–2007), and has been Guest Editor for the European Journal of Control and Control Engineering Practice. He is a Professor at the School of Electrical Engineering, KTH, Stockholm, Sweden. He is a Chair of the IFAC Coordinating Committee CC1 Systems and Signals. In 2001, he received the KTH award for outstanding contribution to undergraduate education. He is co-recipient of the European Research Council advanced grant.