Preprint of final paper. © 2018 IEEE DOI: 10.1109/CDC.2018.8618912

On the identification of linear time invariant systems with guaranteed simulation error bounds

Marco Lauricella and Lorenzo Fagiano

***Abstract*— The model identification problem for asymptotically stable linear time invariant systems is considered. The system output is affected by an additive noise with unknown bound, and a finite set of data is available for parameter estimation. The goal is to derive a model with guaranteed simulation error bounds for all predicted time steps, up to a finite horizon of choice. This is achieved in three steps. At first, the noise bound, system order, and impulse response decay rate are estimated from data. Then, the estimated quantities are used to refine the sets of all possible multi-step predictors compatible with data and prior assumptions (Feasible Parameter Sets, FPSs). The FPSs allow one to derive, in a Set Membership framework, guaranteed error bounds for any given multi-step predictor, including the one obtained by simulating the system model. Finally, the wanted model parameters are identified by numerical optimization, imposing the constraints provided by the FPSs and using one of two proposed optimality criteria. Numerical simulations illustrate the validity of the approach.**

I. Introduction

We investigate the problem of identifying models of linear time invariant systems for simulation purpose. Simulation (or multi-step prediction) is intended as the estimation, by iterating a one-step-ahead model, of the p-steps-ahead future output of the system at hand, where $p \in [1, \bar{p}]$ and $\bar{p}$ is a finite horizon of choice. Models identified by minimizing the one-step-ahead prediction error (Prediction Error Minimization, PEM) do not necessarily provide good estimation performance when applied to the multi-step case, see e.g. [11]. In general, models identified with PEM criterion describe more accurately high-frequency dynamics, while simulation-based methods (Simulation Error Minimization, SEM) capture the low-frequency dynamics [13]. Moreover, we are interested in estimating (and possibly minimizing in the identification procedure) the worst-case bounds on the simulation error between the true system and its model when a known future input signal is applied. Most of the existing approaches in the literature do not consider the problem of deriving such bounds, which are however extremely important in applications such as process analysis and robust control design techniques. In [1], [2], [3], [4], [10], and [14], approaches to identify one-step-ahead models able to predict several steps ahead simultaneously are proposed. The most common approach is to identify a one-step-ahead model by minimizing the prediction or simulation error of the resulting p-step ahead iterated models, for all $p \in [1, \bar{p}]$. These methods are usually cast in a stochastic framework, resulting in stochastic accuracy bounds. Here, we are instead interested in finding worst-case, deterministic bounds under the only assumption of bounded noise (with unknown bound), which is typically satisfied in practical applications. Finally, stability of the identified one-step-ahead predictors is usually not easily guaranteed a priori, as it is difficult to enforce in the identification process.

The estimation method presented in this work aims to address the issues listed above. Our goal is to obtain a model with guaranteed worst-case error bounds in simulation, enforcing at the same time convergence (i.e. stability) properties on its impulse response. This is achieved in three steps. At first, the noise bound, order, and impulse response decay rate of the system at hand are estimated from data. To this end, new results are presented and exploited, pertaining to the estimation of multi-step-ahead linear predictors in a Set Membership framework, see [6], [7], [12]. As a second step, the estimated quantities are used to refine the sets of all possible multi-step models compatible with data and prior assumptions (Feasible Parameter Sets, FPSs). The FPSs are exploited to derive guaranteed error bounds for any given multi-step predictor. Finally, the wanted model parameters are estimated by numerical optimization, imposing the constraints provided by the FPSs and minimizing one of two proposed optimality criteria, namely to minimize the worst-case error bound over the considered simulation horizon, or to minimize a standard simulation error criterion. By adding the estimated decay rate as a further constraint in the FPSs, we can enforce such a trend on the identified model as well, a procedure similar to the one adopted in [8] and [9] for data-driven observer design. Finally, numerical simulations illustrate the validity of the approach and its advantages over standard PEM and SEM methods.

II. Problem formulation

Consider a discrete time, asymptotically stable, strictly proper linear time invariant system of order n, with input $u(k) \in \mathbb{R}$ and output $z(k) \in \mathbb{R}$, where $k \in \mathbb{Z}$ is the discrete time variable. The output measurement $y(k)$ is affected by an additive noise $d(k)$:

$$y(k) = z(k) + d(k),$$

Assumption 1: (Noise and input bounds)

- $|d(k)| \leq \bar{d}_0$, $\forall k \in \mathbb{Z}$.
- $u(k) \in \mathbb{U} \subset \mathbb{R}$, $\forall k \in \mathbb{Z}$, $\mathbb{U}$ compact.

Assumption 2: (Observability and reachability) The system at hand is completely observable and reachable.

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. E-mail addresses: {marco.lauricella | lorenzo.fagiano }@polimi.it. Corresponding author: Marco Lauricella

Since the system is linear, for a given k and any prediction horizon $p \in \mathbb{N}$, it holds:

$$z(k+p) = \underbrace{[Z_n^T(k)\ U_{p,n}^T(k)]}_{\psi_p(k)^T} \underbrace{\begin{bmatrix} \theta_{p,z}^0 \\ \theta_{p,u}^0 \end{bmatrix}}_{\theta_p^0} = \psi_p(k)^T \theta_p^0, \quad (1)$$

where ^{T} denotes the matrix transpose operation, and:

$$\begin{aligned} Z_n(k) &\doteq \begin{bmatrix} z(k) & z(k-1) & \dots & z(k-n+1) \end{bmatrix}^T, \\ U_{p,n}(k) &\doteq [u(k+p-1)\ u(k+p-2)\ \dots\ u(k-n+1)]^T. \end{aligned} \quad (2)$$

For an asymptotically stable system, there exist scalars L_u, L_z and ρ such that:

$$\begin{array}{rcl} |\theta_{p,u}^{0,(i)}| & \leq & L_u \rho^i,\ i = 1, \dots, p+n-1 \\ |\theta_{p,z}^{0,(i)}| & \leq & L_z \rho^{p+i},\ i = 1, \dots, n-1 \end{array}, \quad (3)$$

where $^{(i)}$ denotes the element in the i-th position of a vector. In (3), the decay rate ρ is dictated by the magnitude of the system's dominant poles. Thus, we can write the one-step-ahead dynamics of the true system by considering $p = 1$ in (1):

$$z(k+1) = \psi_1(k)^T \theta_1^0, \quad (4)$$

which corresponds to a standard auto-regressive description with exogenous input (ARX). For any $p > 1$, the entries of the parameter vector θ_p^0 are polynomial functions of the entries of θ_1^0, obtained by recursion of (4). We indicate this polynomial dependency in compact form as:

$$\theta_p^0 = h(\theta_1^0, p, n).$$

As motivated in the introduction, we consider the problem of identifying the parameters of a one-step-ahead model of (4) from data. To this end, we introduce the model regressor $\varphi_p(k) \in \mathbb{R}^{2o+p-1}$, where $o \in \mathbb{N}$ is the chosen model order:

$$\varphi_p(k) \doteq [Y_o^T(k)\ U_{p,o}^T(k)]^T,$$

where $Y_o(k) \doteq [y(k)\, y(k-1) \dots y(k-o+1)]^T \in \mathbb{R}^o$ and $U_{p,o}$ is defined as in (2). Then, we consider the following ARX model structure for our one-step-ahead model:

$$\hat{z}(k+1) = \varphi_1(k)^T \theta_1, \quad (5)$$

where $\hat{z}(k+1) \approx z(k+1)$ is the predicted one-step-ahead output, and $\theta_1 \in \mathbb{R}^{2o}$ is the model parameter vector to be estimated from data. Simulating (i.e. iterating) the model (5) defines the following multi-step predictors for each $p > 1$, $p \in \mathbb{N}$:

$$\hat{z}(k+p) = \varphi_p(k)^T \begin{bmatrix} \hat{\theta}_{p,y} \\ \hat{\theta}_{p,u} \end{bmatrix} = \varphi_p(k)^T \hat{\theta}_p,$$

where $\hat{z}(k+p)$ is the predicted (i.e. simulated) p-step ahead future output, and $\hat{\theta}_p = h(\hat{\theta}_1, p, o) \in \mathbb{R}^{2o+p-1}$ is the corresponding parameter vector, whose entries are polynomial functions of the entries of $\hat{\theta}_1$.

Besides the possible order mismatch (i.e. $o \neq n$), the main difference between the model (5) and the true system (4) is that the former employs noise-affected measurements $y(k)$ of the output in its regressor, instead of the true output values $z(k)$. To study the effects of this difference, let us define the vector $\psi_{p,o}(k) \doteq [Z_o^T(k)\ U_{p,o}^T(k)]^T$, where Z_o is obtained as in (2). Assumption 1, along with the asymptotic stability of the system, implies that the regressors $\psi_{p,o}(k)$ belong to a compact set $\Psi_{p,o}$:

$$\psi_{p,o}(k) \in \Psi_{p,o} \subset \mathbb{R}^{2o+p-1},\ \Psi_{p,o} \text{ compact},\ \forall p \in \mathbb{N},\ \forall k \in \mathbb{Z}.$$

Consequently, $\varphi_p(k)$ belongs to a compact set Φ_p as well:

$$\varphi_p(k) \in \Phi_p = \Psi_{p,o} \oplus \mathbb{D}_p,\ \forall p \in \mathbb{N},\ \forall k \in \mathbb{Z},$$

where $F \oplus M = \{f + m : f \in F,\ m \in M\}$ is the Minkowski sum of two given sets F, M, and

$$\mathbb{D}_p \doteq \{[d^{(1)}, \dots, d^{(o)}, 0, \dots, 0]^T : |d^{(i)}| \leq \bar{d}_0\} \subset \mathbb{R}^{2o+p-1}$$

is the set of all possible noise realizations that can affect the system output values stacked inside φ_p. In practical applications, the sets $\Psi_{p,o}$ and Φ_p depend on the input/output trajectories of the system, and they are not known explicitly. However, for the sake of parameter identification, a finite number N of measured pairs $(\tilde{\varphi}_p(i), \tilde{y}_p(i))$ is assumed to be available, where $\tilde{\cdot}$ denotes a specific sample and $\tilde{y}_p(i) \doteq \tilde{y}(i+p)$. These sampled data define the set:

$$\tilde{\mathscr{V}}_p^N \doteq \left\{ \tilde{v}_p(i) = \begin{bmatrix} \tilde{\varphi}_p(i) \\ \tilde{y}_p(i) \end{bmatrix},\ i = 1, \dots, N \right\} \subset \mathbb{R}^{2o+p}, \quad (6)$$

The continuous counterpart of $\tilde{\mathscr{V}}_p^N$ is:

$$\mathscr{V}_p \doteq \left\{ v_p = \begin{bmatrix} \varphi_p \\ y_p \end{bmatrix} : y_p \in Y_p(\varphi_p),\ \forall \varphi_p \in \Phi_p \right\} \subset \mathbb{R}^{2o+p},$$

where $Y_p(\varphi_p) \subset \mathbb{R}$ is the compact set of all possible measured output values corresponding to every value of $\varphi_p \in \Phi_p$ and every noise realization $d : |d| \leq \bar{d}_0$.

Assumption 3: (Informative content of data) For any $\beta > 0$, there exists a value of $N < \infty$ such that:

$$d_2\left(\mathscr{V}_p, \tilde{\mathscr{V}}_p^N\right) \leq \beta,$$

where $d_2\left(\mathscr{V}_p, \tilde{\mathscr{V}}_p^N\right) \doteq \max_{v_1 \in \mathscr{V}_p} \min_{v_2 \in \tilde{\mathscr{V}}_p^N} \|v_2 - v_1\|_2$.

The meaning of Assumption 3 is that, by adding more points to the measured data-set, the set of all the trajectories of interest is densely covered, leading to $\lim_{N \to \infty} d_2\left(\mathscr{V}_p, \tilde{\mathscr{V}}_p^N\right) = 0$. This corresponds to a persistence of excitation condition, plus a bound-exploring property of the variable $d(k)$.

We can now state the problem addressed in this paper.

Problem 1: Under Assumptions 1-3, use the available data (6) to:

a) estimate the noise bound $\bar{d}_0$, the system order n, and the decay rate ρ;
b) identify the parameters of the model (5) according to a suitable optimality criterion, together with associated guaranteed bounds on the simulation (i.e. multi-step prediction) error $|z(k+p) - \hat{z}(k+p)|$, $p = 1, \dots, \bar{p}$, where $\bar{p} < \infty$ is a maximum simulation horizon of interest.

III. Estimation of noise bound, system order, and decay rate

A. Preliminary results

We start by recalling results derived in [12], which we employ and complement with new ones. Consider a generic $p \in \mathbb{N}$ and a parameter vector θ_p defining a multi-step predictor $\varphi_p(k)^T\theta_p \approx z(k+p)$ (not necessarily computed by iterating a one-step-ahead model). Denoting the error between the system output and such an estimate as $\varepsilon_p(\theta_p, \varphi_p(k)) = z(k+p) - \varphi_p(k)^T\theta_p$, under Assumption 1 we have:

$$\left|y(k+p) - \varphi_p(k)^T\theta_p\right| \leq \bar{\varepsilon}_p(\theta_p) + \bar{d},$$

where $\bar{\varepsilon}_p(\theta_p)$ represents the global error bound produced by θ_p (termed "global" since it holds for all possible regressor values in the set Φ_p), and $\bar{d} \geq 0$ is an estimate of the true noise bound $\bar{d}_0$. $\bar{\varepsilon}_p(\theta_p)$ is given by:

$$\begin{aligned}\bar{\varepsilon}_p(\theta_p) = &\min_{\varepsilon\in\mathbb{R}} \ \varepsilon \ \text{ subject to} \\ &\left|y_p - \varphi_p^T\theta_p\right| \leq \varepsilon + \bar{d}, \ \forall(\varphi_p, y_p) : \begin{bmatrix}\varphi_p \\ y_p\end{bmatrix} \in \mathscr{V}_p\end{aligned}$$

This bound cannot be computed exactly in practice, with a finite set of data points. In [12], a method for estimating $\bar{\varepsilon}_p(\theta_p)$ is proposed, along with the proof that this estimate, denoted with $\underline{\lambda}_p$, converges to $\bar{\varepsilon}_p$ from below. $\underline{\lambda}_p$ is obtained by solving the following linear program (LP):

$$\begin{aligned}\underline{\lambda}_p = &\min_{\theta_p,\lambda\geq 0} \ \lambda \ \text{ subject to} \\ &\left|\tilde{y}_p - \tilde{\varphi}_p^T\theta_p\right| \leq \lambda + \bar{d}, \ \forall(\tilde{\varphi}_p, \tilde{y}_p) : \begin{bmatrix}\tilde{\varphi}_p \\ \tilde{y}_p\end{bmatrix} \in \tilde{\mathscr{V}}_p^N\end{aligned} \tag{7}$$

Then, $\underline{\lambda}_p$ is inflated to account for the uncertainty due to the use of a finite number of measurements:

$$\hat{\bar{\varepsilon}}_p = \alpha\underline{\lambda}_p, \ \alpha > 1. \tag{8}$$

We can now recall the Feasible Parameter Set (FPS) Θ_p, i.e. the set of parameter values that are consistent with the information coming from data and noise bound estimate:

$$\Theta_p = \left\{\theta_p : |\tilde{y}_p - \tilde{\varphi}_p^T\theta_p| \leq \hat{\bar{\varepsilon}}_p + \bar{d}, \ \forall(\tilde{\varphi}_p, \tilde{y}_p) : \begin{bmatrix}\tilde{\varphi}_p \\ \tilde{y}_p\end{bmatrix} \in \tilde{\mathscr{V}}_p^N\right\} \tag{9}$$

If the FPS is bounded, it results in a polytope with at most N faces (if it is unbounded, then the employed data are not informative enough and new data should be collected). The FPS can be used to derive a global bound on the prediction error produced by a given value of θ_p, indicated with $\tau_p(\theta_p)$:

$$\begin{aligned}&|z(k+p) - \hat{z}(k+p)| \leq \tau_p(\theta_p) \\ &\tau_p(\theta_p) = \max_{\varphi_p\in\Phi_p} \max_{\theta\in\Theta_p} |\varphi_p^T(\theta - \theta_p)| + \hat{\bar{\varepsilon}}_p.\end{aligned}$$

Similarly to $\bar{\varepsilon}_p$, also $\tau_p(\theta_p)$ cannot be computed exactly with a finite data set. An estimate is given by:

$$\underline{\tau}_p(\theta_p) = \max_{\tilde{\varphi}_p\in\tilde{\mathscr{V}}_p^N} \max_{\theta\in\Theta_p} |\tilde{\varphi}_p^T(\theta - \theta_p)| + \hat{\bar{\varepsilon}}_p.$$

$\underline{\tau}_p(\theta_p)$ converges to its counterpart $\tau_p(\theta_p)$ from below as N increases under Assumption 3, see [12]. In practical applications, we inflate $\underline{\tau}_p(\theta_p)$ as well, in order to compensate for the uncertainty deriving from the usage of a finite data-set:

$$\hat{\tau}_p(\theta_p) = \gamma\Big(\max_{\tilde{\varphi}_p\in\tilde{\mathscr{V}}_p^N} \max_{\theta\in\Theta_p} \left|\tilde{\varphi}_p^T(\theta - \theta_p)\right|\Big) + \hat{\bar{\varepsilon}}_p, \ \gamma > 1. \tag{10}$$

Assumption 4: (Estimated error bounds) The estimated values of $\hat{\bar{\varepsilon}}_p$ and $\hat{\tau}_p(\theta_p)$ are larger than the corresponding true bounds $\bar{\varepsilon}_p$ and $\tau_p(\theta_p)$, respectively.

In practice, α and γ express how much one is confident on the informative content of the identification experiment. When the data-set is large and informative enough, these scalars can be chosen close to 1. Tuning of α and γ can be carried out by cross-validation.

B. New results on the estimated multi-step error bounds

We present two results showing additional properties of the quantity $\underline{\lambda}_p$ (7). These provide a theoretical justification to the estimation procedures for the noise bound $\bar{d}_0$, system order n, and decay rate ρ, which we propose in Section III-C. Let us define:

$$\lambda_p \doteq \min_{\theta_p\in\Omega} \max_{\left[\begin{smallmatrix}\varphi_p \\ y_p\end{smallmatrix}\right]\in\mathscr{V}_p} \left(\left|y_p - \varphi_p^T\theta_p\right| - \bar{d}\right). \tag{11}$$

In (11), $\Omega \subset \mathbb{R}^{2o+p-1}$ represents a compact approximation of the real set $\mathbb{R}^{2o+p-1}$: it can be chosen e.g. by considering box constraints of $\pm 10^{15}$ on each element of the parameter vector. This is a technical assumption that allows us to use the maximum and minimum operators, instead of supremum and infimum.

Assumption 5: (Model order) The model order o is chosen such that $o \geq n$.

As indicated in Section III-C, this assumption can be satisfied by initially over-estimating the system order, since the results presented below are not affected by the specific value of o, as long as it is larger than n.

Remark 1: With a slight abuse of notation, in the remainder we imply that, when $o \neq n$, the parameter vectors θ_p (if $o < n$), or θ_p^0 (if $o > n$), are appropriately padded with zero entries to equate their dimensions, thus keeping consistency of all matrix operations.

Theorem 1: If Assumptions 1-3 and 5 hold, then:

1) $\lambda_p \xrightarrow{p\to\infty} (\bar{d}_0 - \bar{d})$
2) $\underline{\lambda}_p \leq \lambda_p$
3) $\forall\eta \in (0, \lambda_p], \ \exists N < \infty : \underline{\lambda}_p \geq \lambda_p - \eta$

Proof: See [5]. ∎

Corollary 1: If Assumptions 1-3 and 5 hold, and if the noise bound is correctly chosen as $\bar{d} = \bar{d}_0$, then:

$$\lambda_p = \bar{d}_0\|\theta_{p,z}^0\|_1 \leq n\,\bar{d}_0\,L_z\,\rho^{p+1}$$

Proof: See [5]. ∎

Remark 2: Theorem 1 and Corollary 1 imply two consequences that are useful for model identification. The first is that, when $\bar{d} = \bar{d}_0$ and $o < n$, λ_p converges to a non-zero value as p increases, which is due to the model order

mismatch. The rationale behind this statement is that, when $o < n$, there exists a choice of φ_p, y_p inside $\mathscr{Y}_p$ such that it is not possible to find a θ_p able to bring the error $\varphi_p^T(\theta_p^0 - \theta_p)$ to zero. This observation will be used to estimate the model order in the next section. The second is that $\lambda_p \xrightarrow{p\to\infty} 0$ with the same decay rate as that of the true system parameters, thus providing a way to estimate suahc a decay rate.

C. Estimation of noise bound, system order, and decay rate

From Theorem 1 it follows that, for $N \to \infty$ and $o \geq n$, picking a noise bound estimate $\bar{d} \geq \bar{d}_0$ results in $\underline{\lambda}_p$ converging to zero as p increases; instead, choosing $\bar{d} < \bar{d}_0$ results in $\underline{\lambda}_p$ converging to a non-zero value. We exploit this property to estimate the value of the noise bound $\bar{d}_0$:

Procedure 1 Estimation of $\bar{d}_0$

1) Choose a large value of o as initial guess.
2) Set a starting value of $\bar{d}$ small enough to have $\bar{d} < \bar{d}_0$.
3) Gradually increase $\bar{d}$, recalculating $\underline{\lambda}_p$, until the first value of $\bar{d}$ under which $\exists \bar{p} : \underline{\lambda}_p \simeq 0 \ \forall p > \bar{p}$ is found.
4) The obtained $\bar{d}$ corresponds to the noise bound, and the related $\bar{p}$ represents the system settling time.

Then, exploiting the observation reported in Remark 2, we estimate the model order o:

Procedure 2 Estimation of o

1) Set $\bar{d}$ and $\bar{p}$ to the values resulting from Procedure 1.
2) Choose a large value of o as initial guess.
3) Gradually decrease o, recalculating $\underline{\lambda}_p$, until the first value of o under which $\exists p > \bar{p} : \underline{\lambda}_p > 0$ is found.
4) The last value of o under which $\underline{\lambda}_p \simeq 0 \ \forall p > \bar{p}$ will be the minimal predictor order.

Finally, we exploit the observed decay rate of $\underline{\lambda}_p$ to estimate the quantities $\hat{\rho} \approx \rho$, $\hat{L}_z \approx L_z$, and $\hat{L}_u \approx L_u$, see (3). Let us define $\boldsymbol{f}_{\hat{\varepsilon}} \doteq \begin{bmatrix}\hat{\bar{\varepsilon}}_1 & \cdots & \hat{\bar{\varepsilon}}_{p_{\max}}\end{bmatrix}^T$, where $\hat{\bar{\varepsilon}}_p$ is obtained from (8) with $\bar{d}$ resulting from Procedure 1, and $p_{\max} > \bar{p}$. Let us also define, for given values of L and ρ, the quantities $g_{L\rho}(p) \doteq L\rho^{p+1}$, $p \in [1, p_{\max}]$. Then, we solve the following optimization problem to compute $\hat{\rho}$:

$$\begin{aligned}\left[\hat{L}, \hat{\rho}\right] = &\arg \min_{L,\rho} \ \left\|\boldsymbol{f}_{\hat{\varepsilon}} - \boldsymbol{g}_{L\rho}\right\|_2^2 \ \text{ subject to} \\ &\boldsymbol{g}_{L\rho} \succeq \boldsymbol{f}_{\hat{\varepsilon}} \\ &L > 0, \ 0 < \rho < 1\end{aligned} \tag{12}$$

where $\boldsymbol{g}_{L\rho} = \begin{bmatrix} g_{L\rho}(1) & \cdots & g_{L\rho}(p_{\max})\end{bmatrix}^T$. In practice, the computed value of $\hat{\rho}$ minimizes the quadratic norm of the difference between $\hat{\bar{\varepsilon}}_p$ (i.e. the observed decay rate) and $g_{L\rho}(p)$ (the theoretical exponential decay rate). Supported by Corollary 1, this estimate of $\hat{\rho}$ is consistent with the system decay rate. However, we still need to estimate suitable values of $\hat{L}_z$, $\hat{L}_u$. For the former, we exploit the FPSs Θ_p considering the parameters pertaining to the output values inside the regressors φ_p:

$$\hat{L}_z = \left(\max_{p\in[1,\bar{p}]} \max_{\theta_p\in\Theta_p} \max_{i=1,\ldots,o} \theta_p^{(i)}\right) \Big/ \hat{\rho} \ . \tag{13}$$

Regarding $\hat{L}_u$, we instead consider the parameters pertaining to the o most recent input values inside the regressors φ_p:

$$\hat{L}_u = \left(\max_{p\in[1,\bar{p}]} \max_{\theta_p\in\Theta_p} \max_{i=o+1,\ldots,2o} \theta_p^{(i)}\right) \Big/ \hat{\rho} \ . \tag{14}$$

Indeed, the magnitude of these parameters is not affected by the decay rate and it can be used to estimate the true bounds L_z and L_u (see (3)).

IV. Identification of one-step-ahead predictors with guaranteed simulation error bounds

We are now in position to address part b) of Problem 1. In particular, we present new methods to learn the parameters of one-step-ahead prediction models of the form (5), considering the simulation (multi-step) accuracy and trying to enforce asymptotic stability of the predictor as well. The first step is to refine the FPSs Θ_p (9), by adding constraints that take into account the estimated decay rate. To this end, let us define:

$$\begin{aligned}\Gamma_p = \Big\{\theta_p : &|\theta_{p,u}^{(i)}| \leq \hat{L}_u \hat{\rho}^i, \ \forall i \in [1, p+o-1], \\ &\wedge \ |\theta_{p,y}^{(i)}| \leq \hat{L}_z \hat{\rho}^{p+i}, \ \forall i \in [1, o]\Big\}\end{aligned}$$

Then, we modify the Feasible Parameter Sets as follows:

$$\Theta_p^{L\rho} = \Theta_p \cap \Gamma_p. \tag{15}$$

Assumption 6: (Estimated decay rate) The estimated decay rate parameters are such that $\hat{L}_z \geq L_z$, and $\hat{L}_u \geq L_u$.

Remark 3: Under Assumptions 4, 5 and 6, it follows that $\theta_p^0 \in \Theta_p^{L\rho}, \ \forall p$, i.e. each FPS (15) is non-empty and contains the parameters of the corresponding iterated model of the system (4). These assumptions cannot be verified in practice when a finite data-set is used. However, as long as the sets $\Theta_p^{L\rho}$ are non-empty (which can be easily verified, since they are all polytopes), we can be confident that the computed estimates and prior assumptions are not invalidated by data. Whenever $\Theta_p^{L\rho}$ becomes empty for some p, the estimated bounds can be enlarged until non-empty sets are obtained again.

We describe next two possible procedures to estimate θ_1 (5), exploiting the modified FPSs. Both procedures are based on nonlinear programs.

A. Method I - minimal worst-case simulation error bound

This method is based on the concept of global error bound. We estimate the model that minimizes the maximum worst-case error bound over the whole simulation horizon:

$$\begin{aligned}\hat{\theta}_1 = &\arg \min_{\theta_1} \left\|\boldsymbol{\tau}\left(\boldsymbol{\theta}\right)\right\|_\infty \ \text{ subject to} \\ &\theta_p \in \Theta_p^{L\rho}, \ \forall p \in [1, \bar{p}]\end{aligned}$$

where $\theta_p = h(\theta_1, p, o)$, $\boldsymbol{\tau}\left(\boldsymbol{\theta}\right) = \begin{bmatrix}\hat{\tau}_1(\theta_1) & \cdots & \hat{\tau}_{\bar{p}}(\theta_{\bar{p}})\end{bmatrix}^T$, and $\hat{\tau}_p(\theta_p)$ is defined as in (10). This results in:

$$\begin{aligned}\hat{\theta}_1 = &\arg\min_{\theta_1} \left(\max_{p\in[1,\bar{p}]} \max_{i=1,\ldots,N} \max_{\theta\in\Theta_p^{L\rho}} \left|\tilde{\varphi}_p(i)^T(\theta - \theta_p)\right| + \hat{\bar{\varepsilon}}_p\right) \\ &\text{subject to } \theta_p \in \Theta_p^{L\rho}, \ \forall p \in [1, \bar{p}]\end{aligned} \tag{16}$$

Problem (16) can be rewritten in a simpler form. First, we split the absolute value in two terms:

$$\breve{\varphi}_p(j) = \begin{cases} \tilde{\varphi}_p(j) & \text{if} \quad j \leq N \\ -\tilde{\varphi}_p(j) & \text{if} \quad j > N \end{cases} \quad \text{for} \quad j = 1, \ldots, 2N.$$

Then, by defining $c_{j_p} = \max_{\theta \in \Theta_p^{L\rho}} \breve{\varphi}_p(j)^T \theta$, $j = 1, \ldots, 2N$, $p = 1, \ldots, \bar{p}$, we can reformulate (16) as:

$$\begin{aligned} \hat{\theta}_1 = &\arg \min_{\theta_1} \max_{p \in [1,\bar{p}]} \max_{i=1,\ldots,2N} (c_{j_p} - \breve{\varphi}_p(j)^T \theta_p) \\ &\text{subject to } \theta_p \in \Theta_p^{L\rho}, \ \forall p \in [1, \bar{p}] \end{aligned} \tag{17}$$

The optimization problem defined by (17) corresponds to:

$$\begin{aligned} \hat{\theta}_1 = &\arg \min_{\theta_1, \zeta} \zeta \ \text{ subject to} \\ &c_{j_p} - \breve{\varphi}_p(j)^T \theta_p \leq \zeta, \ j = 1, \ldots, 2N, \ p = 1, \ldots, \bar{p} \\ &\theta_p \in \Theta_p^{L\rho}, \ \forall p \in [1, \bar{p}] \end{aligned}$$

which is a nonlinear program with $2N$ linear constraints and $2N(\bar{p}-1)$ polynomial constraints, plus $2N\bar{p}$ polynomial constraints that require the preliminary solution of $2N\bar{p}$ LP problems.

B. Method II - minimal simulation error on identification data-set

In this method, we minimize the simulation error produced by the one-step iterated prediction model, given a certain initial condition $\varphi_1(0)$, and we enforce the decay rate on the impulse response parameters:

$$\begin{aligned} \hat{\theta}_1 = &\arg \min_{\theta_1 \in \Theta_1^{L\rho}} \left\| \tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Z}}(\theta_1) \right\|_2^2 \ \text{ subject to} \\ &\theta_p \in \Gamma_p, \ \forall p \in [2, \bar{p}] \end{aligned} \tag{18}$$

where $\theta_p = h(\theta_1, p, o)$, $\tilde{\boldsymbol{Y}} = \begin{bmatrix} \tilde{y}(1) & \tilde{y}(2) & \cdots & \tilde{y}(N) \end{bmatrix}^T$, and $\hat{\boldsymbol{Z}}(\theta_1) = \begin{bmatrix} \tilde{\varphi}_1(0)^T \theta_1 & \tilde{\varphi}_2(0)^T \theta_2 & \cdots & \tilde{\varphi}_N(0)^T \theta_N \end{bmatrix}^T$. (18) corresponds to a nonlinear optimization problem, with $2\bar{p}$ polynomial constraints.

V. Simulation results

We test the approach on a third-order system with static gain equal to 10 and complex dominant poles with damping ratio 0.1 in continuous time, whose output is affected by random noise, uniformly distributed in the interval $[-0.1, \ 0.1]$ (i.e. $\bar{d}_0 = 0.1$). Input and output data points are acquired with sampling time $T_s = 0.1$. The data-set collected for the identification phase and the one used for the validation phase contain $N = 1500$ and $N_v = 1500$ samples, respectively. The input signal takes values in the set $\{-1; \ 0; \ 1\}$ randomly every 10 time units. Fig. 1 depicts the behavior of the measured system output during the identification experiment.

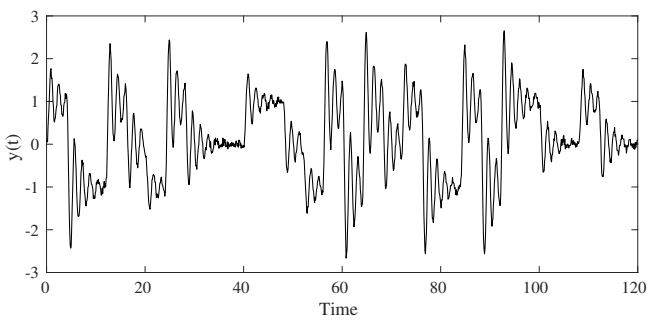

Fig. 1. Measured system output during the identification experiment.

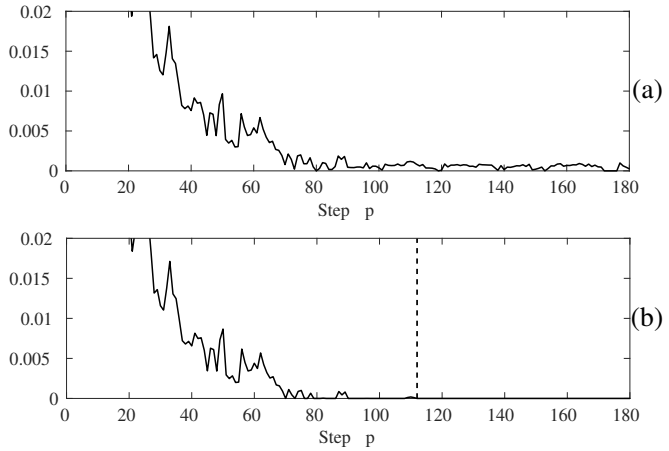

Fig. 2. Estimated values of $\underline{\lambda}_p$ with model order $o = 5$ and different choices of noise bound $\bar{d}$. Fig. (a): $\bar{d} = 0.098$; fig. (b): $\bar{d} = 0.099$. The dashed vertical line indicates the value of $\bar{p}$ corresponding to each choice of $\bar{d}$.

The first step of our identification procedure regards the estimation of the noise bound $\bar{d}$. Adopting Procedure 1, and choosing an initial model order $o = 5$, we calculate $\underline{\lambda}_p$ for several values of $\bar{d}$. Based on the results (depicted in Fig. 2), we decide to set $\bar{d} = 0.099$, with corresponding settling time $\bar{p} = 115$. These values are actually consistent with the true ones, since $\bar{d}_0 = 0.1$ and the time constant corresponding to the dominant poles of $G(s)$ is $T = 2.5$, resulting in a settling time of about 125 steps. Then, we apply Procedure 2 to select the model order. The obtained result is $o = 3$, which is consistent with the (a priori unknown) order of the considered system and verifies Assumption 5.

Having defined our choice of $\bar{d}$ and o, we perform the procedure proposed in Section III-C to estimate the system decay rate. Here $\hat{\rho}$ is estimated as in (12); then, $\hat{L}_z$ and $\hat{L}_u$ are chosen as in (13) and (14), respectively. This results in $\hat{L}_z = 1.871$, $\hat{L}_u = 0.679$ and $\hat{\rho} = 0.965$. For a comparison, the true system's decay rate is $\rho = 0.96$.

Then, the values of $\underline{\lambda}_p$ corresponding to the chosen order o and noise bound $\bar{d}$ are inflated with a coefficient $\alpha = 1.3$, as motivated in Section III. Finally, we set $\gamma = 1.2$. The resulting $\hat{\bar{\varepsilon}}_p$ are used alongside $\bar{d}$, $\hat{L}_z$, $\hat{L}_u$ and $\hat{\rho}$ to define the FPSs for all $p \in [1, \bar{p}]$, as in (15).

A more detailed description of the simulation results and of the estimation procedures can be found in [5].

Finally, we adopt the identification approaches presented in Section IV to estimate the model parameters, and calculate the related guaranteed accuracy bounds as in (10).

As benchmarks for the proposed identification approaches, we consider a one-step-ahead prediction model identified

according to the classical PEM criterion, another one identified using the SEM criterion, and the decoupled multi-step models, identified as proposed in [12]. Each of these decoupled multi-step models is the one that minimizes the corresponding global error bound $\hat{\tau}_p(\theta_p^*)$, thus providing the best performance achievable for every step p in terms of minimization of the guaranteed error bound.

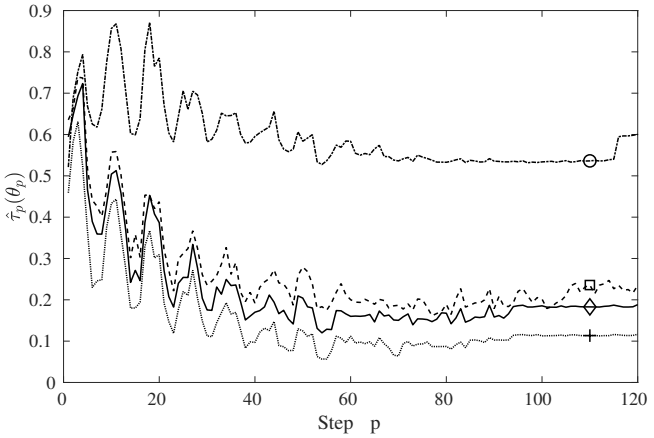

Fig. 3. Guaranteed worst-case error bound. Dotted line with '+': multi-step approach; solid line with '◇': Method II; dashed line with '□': SEM approach; dash-dot line with '○': PEM approach.

We use as performance indicators the guaranteed error bounds and the validation errors produced by each identification approach. The validation error for the p-step ahead model, calculated over the validation data-set, is defined as $e_p = \min_{k=o,\ldots,N_v-p} |z(k+p) - \hat{z}(k+p)|$. Fig. 3 depicts the behavior of the guaranteed error bound corresponding to the various identification methods. Table I presents the values of the worst-case error bound and of the validation error of the p-step ahead model, for some values of p.

TABLE I

COMPARISON BETWEEN VALUES OF $\hat{\tau}_p$ AND e_p OBTAINED BY THE PROPOSED IDENTIFICATION METHODS AND THE BENCHMARK MODELS.

	PEM		SEM		Method I		Method II		Multi-step	
	$\hat{\tau}_p$	e_p	$\hat{\tau}_p$	e_p	$\hat{\tau}_p$	e_p	$\hat{\tau}_p$	e_p	$\hat{\tau}_p$	e_p
$p=1$	0.521	0.199	0.636	0.211	0.531	**0.186**	0.594	0.195	0.459	0.184
$p=10$	0.857	0.367	0.557	0.197	**0.536**	0.163	0.504	**0.158**	0.433	0.193
$p=35$	0.646	0.412	0.262	0.114	0.234	**0.076**	0.235	0.078	0.166	0.081
$p=115$	0.540	0.414	0.227	0.076	0.185	**0.053**	0.187	0.074	0.116	0.083

The presented numerical results show that the proposed approaches obtain better performances in terms of guaranteed error bound and validation error, with respect to both the classic PEM and SEM approaches. In particular, Method II (Section IV-B), which is based on the simulation error cost, is able to significantly improve the performance (both worst-case and actual error with validation data) of the SEM estimation approach without increasing excessively the complexity of the optimization problem.

VI. CONCLUSIONS

We presented new methods to learn one-step-ahead prediction models that provide guaranteed and minimal simulation error bounds. We resorted to the Set Membership identification framework to evaluate and optimize the worst-case simulation error, and presented new results pertaining to the estimation of noise bound, system order, and decay rate. These estimates are then employed to enforce a converging behavior also to the identified model. Finally, we proposed two possible methods to identify the model, and compared them with standard PEM and SEM approaches by means of numerical simulations. The main outcome of the presented work is that the new approaches are able to improve over standard SEM methods, in terms of both guaranteed error bounds and actual accuracy with validation data. In one of the proposed approaches, this comes with minor additional computational complexity. Future work will be devoted to prove additional theoretical properties of the proposed identification approach.

REFERENCES

[1] M. Farina and L. Piroddi. Simulation error minimization identification based on multi-stage prediction. *International Journal of Adaptive Control and Signal Processing*, 25(5):389–406, 2011.

[2] F. Fusco and J.V. Ringwood. Short-term wave forecasting with ar models in real-time optimal control of wave energy converters. In *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, pages 2475–2480. IEEE, 2010.

[3] R.B. Gopaluni, R.S. Patwardhan, and S.L. Shah. MPC relevant identificationtuning the noise model. *Journal of Process Control*, 14(6):699 – 714, 2004.

[4] R. Haber, U. Schmitz, and R. Bars. Long-range optimal model and multi-step-ahead prediction identification for predictive control. *IFAC Proceedings Volumes*, 36(16):477–482, 2003.

[5] M. Lauricella and L. Fagiano. Guaranteed simulation error bounds for linear time invariant systems identified from data. *arXiv:1809.01061*, available at https://arxiv.org/abs/1809.01061, 2018.

[6] M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter. *Bounding approaches to system identification*. Springer Science & Business Media, 2013.

[7] M. Milanese and C. Novara. Model quality in identification of nonlinear systems. *IEEE Transactions on Automatic Control*, 50(10):1606–1611, 2005.

[8] M. Milanese, F. Ruiz, and M. Taragna. Direct data-driven filter design for uncertain LTI systems with bounded noise. *Automatica*, 46(11):1773–1784, 2010.

[9] C. Novara, F. Ruiz, and M. Milanese. Direct filtering: A new approach to optimal filter design for nonlinear systems. *IEEE Transactions on Automatic Control*, 58(1):86–99, 2013.

[10] D.S. Shook, C. Mohtadi, and S.L. Shah. A control-relevant identification strategy for GPC. *IEEE Transactions on Automatic Control*, 37(7):975–980, 1992.

[11] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.

[12] E. Terzi, L. Fagiano, M. Farina, and R. Scattolini. Learning multi-step prediction models for receding horizon control. In *Proc. of 17th European Control Conference*, 2018, pre-print available at http://home.deib.polimi.it/fagiano/docs/TR-TFFS-ECC18a.pdf.

[13] B. Wahlberg and L. Ljung. Design variables for bias distribution in transfer function estimation. *IEEE Transactions on Automatic Control*, 31(2):134–144, 1986.

[14] E. Žáčeková, S. Privara, and Z. Váňa. Model predictive control relevant identification using partial least squares for building modeling. In *Australian Control Conference (AUCC), 2011*, pages 422–427. IEEE, 2011.