An Iterative Identification Method for Linear Continuous-Time Systems

Marco C. Campi, *Senior Member, IEEE*, Toshiharu Sugie, *Fellow, IEEE*, and Fumitoshi Sakai, *Member, IEEE*

***Abstract*—This paper presents a novel approach for the identification of continuous-time systems directly from sampled I/O data based on trial iterations. The method achieves identification through iterative learning control (ILC) concepts in the presence of heavy measurement noise. The robustness against measurement noise is achieved through 1) projection of continuous-time I/O signals onto a finite dimensional parameter space and 2) Kalman filter type noise reduction. In addition, an alternative simpler method is given with some robustness analysis. The effectiveness of the method is demonstrated through numerical examples for a nonminimum phase plant.**

***Index Terms*—Continuous-time systems, iterative learning control, Kalman filter, system identification.**

I. Introduction

ONE of the most important issues in control system design is to obtain an accurate model of the plant to be controlled. Though most of the existing identification methods are described in discrete-time, it would often be convenient to have continuous-time models directly from the sampled I/O data. Indeed many controller design approaches are cast in a continuous-time set-up and, moreover, it is often times easier for us to capture the plant dynamics intuitively in continuous-time rather than in discrete-time.

A basic difficulty of continuous-time identification is that standard approaches (at times called *direct* methods) require to compute the time-derivatives of I/O data, a nontrivial and very delicate task in the presence of measurement noise. A comprehensive survey of this difficulty and of the attempts made to overcome it has been first given by [19] and then by [17]. For more information on direct methods, the reader is referred to the book [14]. Furthermore, the Continuous-Time System Identification (CONTSID) tool-box has been developed on the basis of these direct methods [5]–[7].

On a different topic, iterative learning control (ILC) has attracted much attention over the last two decades as a powerful model-free control methodology [1]–[4], [10], [11], [18]. ILC returns the input which achieves output tracking by iteration of trials for uncertain systems. Though ILC can deal with plants having large uncertainty, most ILC approaches need time-derivatives of I/O data in the continuous-time case [15], and therefore it is quite sensitive to measurement noise. Recently, Hamamoto and Sugie [8], [9] proposed an ILC where the learning law works in a certain finite-dimensional subspace and showed that time-derivatives of the tracking error is not required to achieve perfect tracking in the proposed scheme. Based on this work, Sugie and Sakai [13], [16] proposed an ILC which works in the presence of heavy measurement noise (more than 30% noise to signal ratio) and, moreover, the method was shown to be applicable to the identification (as opposed to the control) of continuous-time systems as well. This identification method proved several advantages such as: 1) no time-derivatives of I/O data are required, 2) it delivers unbiased estimations, and 3) the identified model quality can be estimated by inspecting the tracking performance through experiments. An important restriction applies to this method though: only models with no zeros can be dealt with and this restriction should be removed.

The purpose of this paper is to introduce an identification method for linear continuous-time systems with poles and zeros based on the ILC approach proposed in [13] and [16]. An additional important advantage of the method proposed herein is that, differently from the algorithm in [13] and [16], it guarantees zero convergence of the parameter estimation error as the number of trials increases.

The following notations will be used. Superscript denotes the trial number and subscript denotes the element of a set or a matrix. For instance, input u at the kth trial is written as u^k while $\boldsymbol{x}_i$ is the ith element of the vector $\boldsymbol{x}$.

Section II contains the system description, while the new identification procedure is given in Section III. Section IV presents a simplified scheme and implementation issues are discussed in Section V. Finally, simulation examples are given in Section VI.

II. System Description

Consider the continuous-time SISO system described by

$$y(t) = \frac{B^\circ(p)}{A^\circ(p)}u(t) \triangleq \frac{\beta_0^\circ + \beta_1^\circ p + \cdots + \beta_m^\circ p^m}{1 + \alpha_1^\circ p + \cdots + \alpha_n^\circ p^n}u(t)$$

where $u(t)$ and $y(t), t \in [0, T]$ are the input and the output, respectively, $\alpha_i^\circ \in \mathbb{R}(i = 1, \ldots, n)$ and $\beta_i^\circ \in \mathbb{R}(i = 0, 1, \ldots, m)$ are coefficient parameters, while p is the differential operator, i.e., $pu(t) = \mathrm{d}u(t)/\mathrm{d}t$. We assume the following.

- Many experiments on the system can be repeated and the system can be set any time at rest, that is, it can be operated with zero initial state.
- Though the true parameters α_i° and β_i° are unknown, $A^\circ(p)$ and $B^\circ(p)$ are coprime and their order n and m are known.

Manuscript received November 1, 2006; revised May 17, 2007. Current version published September 10, 2008. This work was supported by MIUR project New Methods for Identification and Adaptive Control for Industrial Systems. Recommended by Associate Editor Brett Ninness.

M. C. Campi is with the Department of Electronics for Automation, University of Brescia, 25123 Brescia, Italy (e-mail: marco.campi@ing.unibs.it).

T. Sugie is with the Department of Systems Science, Graduate School of Informatics, Kyoto University Uji, Kyoto 611-0011, Japan (e-mail: sugie@i.kyoto-u.ac.jp).

F. Sakai is with the Department of Mechanical Engineering, Nara National College of Technology Yamatokoriyama, Nara 639-1080, Japan (e-mail: sakai@mech.nara-k.ac.jp).

Digital Object Identifier 10.1109/TAC.2008.929371

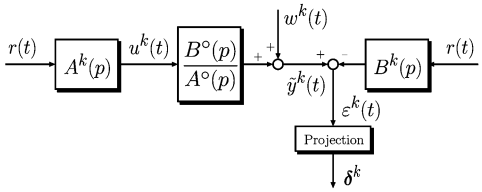

Fig. 1. Data generation scheme at kth trial.

- We can measure $\tilde{y}(t)$, the output contaminated with noise

$$\tilde{y}(t) = y(t) + w(t)$$

where $w(t)$ is zero-mean measurement noise.

The goal is to determine a model in the class

$$\mathcal{M} = \left\{ \frac{B(p)}{A(p)} = \frac{\beta_0 + \beta_1 p + \cdots + \beta_m p^m}{1 + \alpha_1 p + \cdots \alpha_n p^n} \right\}$$

based on I/O measurements $u(t)$ and $\tilde{y}(t)$.

III. Identification Procedure

A. Data Generation Scheme and Parameter Updating Law

Choose a smooth signal $r(t), t \in [0,T]$, such that $r(0) = 0, \frac{\mathrm{d}}{\mathrm{d}t}r(0) = 0, \ldots, \frac{\mathrm{d}^{n+m}}{\mathrm{d}t^{n+m}}r(0) = 0$. At the kth trial, perform the following experiment which produces the signal $\varepsilon^k(t)$ when the parameter estimates $\alpha_1^k, \ldots, \alpha_n^k, \beta_0^k, \ldots, \beta_m^k$ from the previous trial are given (see Fig. 1).

1) Define
$$A^k(p) = 1 + \alpha_1^k p + \cdots + \alpha_n^k p^n$$
$$B^k(p) = \beta_0^k + \beta_1^k p + \cdots + \beta_m^k p^m.$$
2) Compute $u^k(t) = A^k(p)r(t)$.
3) Inject $u^k(t)$ into the system, and collect $\tilde{y}^k(t)$.
4) Compute $B^k(p)r(t)$.
5) Compute the mismatch signal $\varepsilon^k(t)$ by $\varepsilon^k(t) = \tilde{y}^k(t) - B^k(p)r(t)$.

Note that $\varepsilon^k(t) = w^k(t)$ if $A^\circ(p) = A(p)$ and $B^\circ(p) = B(p)$, that is the effect of $r(t)$ on $\varepsilon^k(t)$ disappears. Moreover, $\varepsilon^k(t)$ is obtained without taking any derivative of noisy measurements, only derivatives of $r(t)$ are required. Note also that $\varepsilon^k(t)$ can also be written as

$$\begin{aligned}\varepsilon^k(t) &= \tilde{y}^k(t) - B^k(p)r(t) \\ &= \left(\frac{B^\circ(p)}{A^\circ(p)}A^k(p)r(t) + w^k(t)\right) - B^k(p)r(t) \\ &= \frac{B^\circ(p)A^k(p) - B^k(p)A^\circ(p)}{A^\circ(p)}r(t) + w^k(t).\end{aligned}$$

Now, we introduce $n+m+1$ square-integrable functions $f_1(t), \ldots, f_{n+m+1}(t), t \in [0,T]$, which satisfy the following condition.

Condition 1: If

$$\int_0^T \left(\frac{B^\circ(p)A(p) - B(p)A^\circ(p)}{A^\circ(p)}r(t)\right) \cdot f_i(t)\mathrm{d}t = 0$$
$$i = 1, \ldots, n+m+1 \qquad (1)$$

are satisfied, then $B^\circ(p)A(p) - B(p)A^\circ(p) = 0$ holds. $*$

In words, Condition 1 requires that if the projection of $\frac{B^\circ(p)A(p)-B(p)A^\circ(p)}{A^\circ(p)}r(t)$ onto $f_i(t)$ is zero for all i's, then $\frac{B^\circ(p)A(p)-B(p)A^\circ(p)}{A^\circ(p)}r(t)$ must be zero. Since $\frac{B^\circ(p)A(p)-B(p)A^\circ(p)}{A^\circ(p)}r(t)$ linearly depends on $n+m+1$ parameters, determining $f_i(t)$'s such that this condition is satisfied is not difficult.

Remark 1: One sensible choice for $r(t)$ and for functions $f_i(t)$ corresponds to select a "well-exciting" signal $g(t)$ and to let $r(t) = \bar{A}(p)g(t), f_1(t) = g(t), f_2(t) = \frac{\mathrm{d}}{\mathrm{d}t}g(t), \ldots, f_{n+m+1}(t) = \frac{\mathrm{d}^{n+m}}{\mathrm{d}t^{n+m}}g(t)$, where $\bar{A}(p)$ is some estimate (either *a priori* or obtained during the identification process) of $A^\circ(p)$. If we assume for simplicity that $\bar{A}(p) = A^\circ(p)$, then $\varepsilon^k(t) = (B^\circ(p)A^k(p) - B^k(p)A^\circ(p))g(t) + w^k(t)$. The "useful" part of $\varepsilon^k(t)$, i.e., $(B^\circ(p)A^k(p) - B^k(p)A^\circ(p))g(t)$, is signal $g(t)$ filtered through a polynomial in p of order $n+m$, so that all useful information lies in the subspace generated by $g(t), \frac{\mathrm{d}}{\mathrm{d}t}g(t), \ldots, \frac{\mathrm{d}^{n+m}}{\mathrm{d}t^{n+m}}g(t)$, namely by $f_1(t), f_2(t), \ldots, f_{n+m+1}(t)$, and projecting $\varepsilon^k(t)$ onto this subspace preserves the information while filtering out noise. Another choice is to take $r(t)$ to be a multi-sinusoidal signal with $n+m+1$ components and let each $f_i(t)$ be one of the sinusoids. $*$

Note also that $B^\circ(p)A(p) - B(p)A^\circ(p) = 0$ implies that $B(p) = B^\circ(p)$ and $A(p) = A^\circ(p)$ since $A^\circ(p)$ and $B^\circ(p)$ are coprime.

We next regard $\varepsilon^k(t)$ and $f_1(t), f_2(t), \ldots, f_{n+m+1}(t)$ as elements of $L_2[0,T]$, the space of square-integrable functions defined over $[0,T]$ with inner product $\langle \xi(t), \eta(t) \rangle \triangleq \int_0^T \xi(t)\eta(t)\mathrm{d}t$, and project $\varepsilon^k(t)$ onto the finite-dimensional subspace described by

$$\mathcal{F} \triangleq \mathrm{span}\{f_1(t), f_2(t), \ldots, f_{n+m+1}(t)\}.$$

The projection is written as

$$\varepsilon^k(t)|_{\mathcal{F}} = \delta_1^k f_1(t) + \cdots + \delta_{n+m+1}^k f_{n+m+1}(t) \qquad (2)$$

and $\boldsymbol{\delta}^k \triangleq [\delta_1^k, \ldots, \delta_{n+m+1}^k]^T$ is its vector representation.

We are now in a position to describe how parameters $\alpha_1^k, \ldots, \alpha_n^k, \beta_0^k, \ldots, \beta_m^k$ are updated. Let for brevity

$$\boldsymbol{\gamma}^\circ = [\alpha_1^\circ, \ldots, \alpha_n^\circ, \beta_0^\circ, \ldots, \beta_m^\circ]^T$$
$$\boldsymbol{\gamma}^k = \left[\alpha_1^k, \ldots, \alpha_n^k, \beta_0^k, \ldots, \beta_m^k\right]^T.$$

Then the iterative identification procedure is described as follows (where $\epsilon_0 > 0$ is chosen by the designer).

Step 0) Fix an initial estimate $\boldsymbol{\gamma}^0$, set $k = 0$.

Step 1) Generate $\boldsymbol{\delta}^k$ from $\boldsymbol{\gamma}^k$ according to the scheme shown in Fig. 1.

Step 2) Update $\boldsymbol{\gamma}^k$ with the following rule:

$$\boldsymbol{\gamma}^{k+1} = \boldsymbol{\gamma}^k + H^k\boldsymbol{\delta}^k \qquad (3)$$

where H^k is a learning gain. If $\|\boldsymbol{\gamma}^{k+1} - \boldsymbol{\gamma}^k\| \le \epsilon_0$, stop the iterations. Otherwise, set $k = k+1$ and go back to Step 1.

The choice of H^k will be discussed in the next subsection.

B. Optimal Selection of H^k

First, we rewrite $\boldsymbol{\delta}^k$ in a more convenient form.

Let for the time being $w^k(t) = 0$ [$w^k(t)$ will be reintroduced later]. A simple inspection reveals that step 1 in the identifica-

tion procedure defines an affine operator from $\mathbb{R}^{n+m+1}$ (space for $\boldsymbol{\gamma}^k$) to $\mathbb{R}^{n+m+1}$ (space for $\boldsymbol{\delta}^k$), that is

$$\boldsymbol{\delta}^k = M\boldsymbol{\gamma}^k + \bar{\boldsymbol{\delta}} \tag{4}$$

where M is an $(n+m+1)\times(n+m+1)$ matrix (depending on the true system) and $\bar{\boldsymbol{\delta}}$ is the offset term. Note that from Condition 1 it follows that $\boldsymbol{\delta}^k = 0$ implies $\boldsymbol{\gamma}^k = \boldsymbol{\gamma}^\circ$. This also means that the equation

$$0 = M\boldsymbol{\gamma}^k + \bar{\boldsymbol{\delta}}$$

has the only solution $\boldsymbol{\gamma}^k = \boldsymbol{\gamma}^\circ$, so that M is nonsingular and $\bar{\boldsymbol{\delta}} = -M\boldsymbol{\gamma}^\circ$. Thus, (4) can be rewritten as

$$\boldsymbol{\delta}^k = M(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)$$

with M nonsingular.

When noise $w^k(t)$ is taken into account, $\boldsymbol{\delta}^k$ becomes

$$\boldsymbol{\delta}^k = M(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ) + \boldsymbol{\nu}^k \tag{5}$$

where $\boldsymbol{\nu}^k$ accounts for the projection of $w^k(t)$ onto $\mathcal{F}$.

Inserting (5) in (3) yields

$$\boldsymbol{\gamma}^{k+1} = \boldsymbol{\gamma}^k + H^k M(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ) + H^k\boldsymbol{\nu}^k.$$

Thus, defining $\tilde{\boldsymbol{\gamma}}^k \triangleq \boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ$, we have

$$\tilde{\boldsymbol{\gamma}}^{k+1} = (I + H^k M)\tilde{\boldsymbol{\gamma}}^k + H^k\boldsymbol{\nu}^k \tag{6}$$

which is the equation that describes how the error $\tilde{\boldsymbol{\gamma}}^k$ propagates through trials.

Now, define

$$N \triangleq E[\boldsymbol{\nu}^k(\boldsymbol{\nu}^k)^T], \quad P^k \triangleq E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T]$$

(we assume for simplicity that N is constant through trials).

We next discuss how to select H^k so as to reduce P^k optimally under the assumption that M and N are known. The obtained results will drive us later in the selection of H^k when this assumption will be relaxed. The computation to come are in line with Kalman filtering variance minimization. Noise is assumed to be independent in different experiments.

From (6), we have

$$\begin{aligned} P^{k+1} &= E[\{(I + H^k M)\tilde{\boldsymbol{\gamma}}^k + H^k\boldsymbol{\nu}^k\} \\ &\quad \times \{(I + H^k M)\tilde{\boldsymbol{\gamma}}^k + H^k\boldsymbol{\nu}^k\}^T] \\ &= (I + H^k M)P^k(I + H^k M)^T + H^k N(H^k)^T. \end{aligned} \tag{7}$$

Therefore, P^{k+1} is minimized by the choice

$$H^k = -P^k M^T(MP^k M^T + N)^{-1}. \tag{8}$$

With this choice, we obtain

$$P^{k+1} = P^k - P^k M^T(MP^k M^T + N)^{-1}MP^k. \tag{9}$$

Equations (8) and (9), where (9) is initialized with $P^0 = E[\tilde{\boldsymbol{\gamma}}^0(\tilde{\boldsymbol{\gamma}}^0)^T]$, give the way to select H^k.

The following theorem proves that the proposed method gives us the true parameter $\boldsymbol{\gamma}^\circ$ in the presence of measurement noise through iteration of trials.

Theorem 1: With the updating law (3) where H^k is given by (8) and (9), and (9) is initialized with $P^0 = E[\tilde{\boldsymbol{\gamma}}^0(\tilde{\boldsymbol{\gamma}}^0)^T]$, it holds that

$$P^k = E[(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)^T] \to 0, \text{ as } k \to \infty.$$

Proof: In (7), take $H^k = -\rho M^{-1}$ with $0 < \rho < 1$ [instead of (8)] and denote by P_ρ^k the solution of (7) obtained with initialization $P^0 = E[\tilde{\boldsymbol{\gamma}}^0(\tilde{\boldsymbol{\gamma}}^0)^T]$. We have

$$P_\rho^{k+1} = (1-\rho)^2 P_\rho^k + M^{-1}N(M^{-1})^T\rho^2$$

from which

$$\begin{aligned} P_\rho^k &\le (1-\rho)^{2k}P^0 + M^{-1}N(M^{-1})^T\rho^2\sum_{n=0}^{\infty}(1-\rho)^{2n} \\ &= (1-\rho)^{2k}P^0 + M^{-1}N(M^{-1})^T\frac{\rho}{2-\rho} \end{aligned}$$

holds and the first term vanishes as k grows.

Now, we claim that $P^k \le P_\rho^k$ holds for any k. By induction $P^0 = P_\rho^0$. Assume $P^k \le P_\rho^k$, then we obtain

$$\begin{aligned} P^{k+1} &= (I + H^k M)P^k(I + H^k M)^T + H^k N(H^k)^T \\ &\le [\text{since } H^k \text{in (8) is optimal}] \\ &\le (I - \rho M^{-1}M)P^k(I - \rho M^{-1}M)^T \\ &\quad + \rho M^{-1}N\rho(M^{-1})^T \\ &= (1-\rho)^2 P^k + M^{-1}N(M^{-1})^T\rho^2 \\ &\le (1-\rho)^2 P_\rho^k + M^{-1}N(M^{-1})^T\rho^2 \\ &= P_\rho^{k+1} \end{aligned}$$

so closing the induction. Consequently, we have

$$\lim_{k\to\infty} P^k \le \lim_{k\to\infty} P_\rho^k = M^{-1}N(M^{-1})^T\frac{\rho}{2-\rho}.$$

The right-hand side can be made arbitrarily small by selecting ρ close to zero while the left-hand-side does not depend on ρ, so proving that the left-hand side is actually zero. $*$

It is easy to see that the convergence result that $E[(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)^T] \to 0$ is preserved if $E[\tilde{\boldsymbol{\gamma}}^0(\tilde{\boldsymbol{\gamma}}^0)^T]$ is not known and (9) is instead initialized with a conventional $P^0 > 0$.

The gain determined by (8) and (9) is optimal and should therefore be implemented as such when M and N are known or when a good estimate of M and N is available. Such a gain performs an optimal compromise between exploitation of information and rejection of noise. In the case when $N = 0$ (no noise), we have

$$H^0 = -P^0 M^T(MP^0 M^T + 0)^{-1} = -M^{-1}$$

so that

$$\tilde{\boldsymbol{\gamma}}^1 = (I + H^0 M)\tilde{\boldsymbol{\gamma}}^0 + H^0\boldsymbol{\nu}^0 = 0 + 0 = 0$$

and the error goes to zero in one step.

IV. Simplified Learning Gain

Simpler updating rules for H^k than that given by (8) and (9) can be designed that still guarantee the error convergence to zero. One such rule is given by

$$H^k = -\frac{1}{k+1}M^{-1}. \tag{10}$$

A. Convergence Analysis

Theorem 2: If the updating law (3) is adopted where H^k is given by (10), then

$$E[(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)^T] \to 0, \quad \text{as } k \to \infty.$$

Proof: Since

$$\begin{aligned}\tilde{\boldsymbol{\gamma}}^{k+1} &= (I + H^k M)\tilde{\boldsymbol{\gamma}}^k + H^k \boldsymbol{\nu}^k \\ &= \left(I - \frac{1}{k+1}M^{-1}M\right)\tilde{\boldsymbol{\gamma}}^k - \frac{1}{k+1}M^{-1}\boldsymbol{\nu}^k \\ &= \frac{k}{k+1}\tilde{\boldsymbol{\gamma}}^k - \frac{1}{k+1}M^{-1}\boldsymbol{\nu}^k\end{aligned}$$

holds, we have

$$\begin{aligned}&E[\tilde{\boldsymbol{\gamma}}^{k+1}(\tilde{\boldsymbol{\gamma}}^{k+1})^T] \\ &\quad= \frac{k^2}{(k+1)^2}E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T] + \frac{1}{(k+1)^2}M^{-1}N(M^{-1})^T \\ &\quad= \frac{1}{(k+1)^2}M^{-1}N(M^{-1})^T \\ &\qquad+ \frac{k^2}{(k+1)^2}\cdot\frac{1}{k^2}M^{-1}N(M^{-1})^T \\ &\qquad+ \cdots \\ &\qquad+ \frac{k^2}{(k+1)^2}\cdot\frac{(k-1)^2}{k^2}\cdots\frac{1}{1^2}M^{-1}N(M^{-1})^T \\ &\quad= \underbrace{\left[\frac{1}{(k+1)^2} + \cdots + \frac{1}{(k+1)^2}\right]}_{(k+1)-\text{times}} M^{-1}N(M^{-1})^T \\ &\quad= \frac{1}{k+1}M^{-1}N(M^{-1})^T. \end{aligned} \tag{11}$$

The above term vanishes as $k \to \infty$, which proves the theorem. $*$

Remark 2: One important merit of (10) is stated as follows. From (11), it is obvious that if $N = 0$ (no noise), we get $E[\tilde{\boldsymbol{\gamma}}^1(\tilde{\boldsymbol{\gamma}}^1)^T] = 0$, i.e., the error goes to zero in one step like for the optimal choice (8) and (9). The reason why this happens is that $H^0 = -M^{-1}$ so that $\tilde{\boldsymbol{\gamma}}^1 = (I - M^{-1}M)\boldsymbol{\gamma}^0 - M^{-1}\boldsymbol{\nu}^0 = -M^{-1}\boldsymbol{\nu}^0$ and the initial error is killed in 1 step. In general, though not null, we expect that N is small since $\boldsymbol{\nu}^k$ just represents the noise component in the $(n+m+1)$-dimensional subspace $\mathcal{F}$. $*$

B. Robustness Analysis

Suppose now we only have an estimate $\hat{M}$ of M, so that we implement

$$H^k = -\frac{1}{k+1}\hat{M}^{-1} \tag{12}$$

in place of (10). We next want to investigate conditions under which $E[(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)^T] \to 0$ still holds by using (12).

Theorem 3: If $x^T\hat{M}^{-1}Mx > 0$ holds for $\forall x \neq 0$, then $E[(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)^T] \to 0$ with the choice in (12).

Remark 3: $x^T\hat{M}^{-1}Mx > 0$ is a sign condition and in the scalar case it reduces to say that $\hat{M}$ has the same sign as M. This appears to be a minimal requirement since, if no sign knowledge is available, information δ^k cannot be used to improve the estimate. We thus see that (12) provides an updating gain leading to asymptotic consistency under mild assumptions. $*$

Proof: Consider function $x^T\hat{M}^{-1}Mx$ restricted to the unit sphere, that is, for x such that $\|x\| = 1$. Since the unit sphere is compact and function $x^T\hat{M}^{-1}Mx$ is continuous, the $\inf$ of this function over the unit sphere coincides with its $\min$, which, in turn, has to be positive by the assumption that $x^T\hat{M}^{-1}Mx > 0, \forall x \neq 0$. Thus, there exists a $\mu > 0$ (which we also take to be less than 1 for convenience) such that

$$x^T\hat{M}^{-1}Mx \geq \mu, \quad \forall x : \|x\| = 1.$$

Hence, for any x with $\|x\| = 1$, we obtain

$$\begin{aligned}&\|(I + H^k M)x\|^2 \\ &\quad= \left\|\left(I - \frac{1}{k+1}\hat{M}^{-1}M\right)x\right\|^2 \\ &\quad= x^T\left(I - \frac{1}{k+1}M^T(\hat{M}^{-1})^T\right) \\ &\qquad\times\left(I - \frac{1}{k+1}\hat{M}^{-1}M\right)x \\ &\quad= 1 + \frac{1}{(k+1)^2}\|\hat{M}^{-1}Mx\|^2 - \frac{2}{k+1}x^T\hat{M}^{-1}Mx \\ &\quad\leq \left(1 - \frac{2\mu}{k+1}\right) + \frac{1}{(k+1)^2}\|\hat{M}^{-1}M\|^2 \\ &\quad\leq [\text{for any } k \text{ large enough, say } k \geq \bar{k}] \\ &\quad\leq 1 - \frac{\mu}{k+1}\end{aligned}$$

which implies

$$\|I + H^k M\|^2 \leq 1 - \frac{\mu}{k+1}, \quad \forall k \geq \bar{k}. \tag{13}$$

Consider now (6) and solve it from time $k = \bar{k}$ onward to obtain

$$\tilde{\boldsymbol{\gamma}}^k = \prod_{j=\bar{k}}^{k-1}(I + H^j M)\tilde{\boldsymbol{\gamma}}^{\bar{k}} + \sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}(I + H^j M)\cdot H^r\boldsymbol{\nu}^r \tag{14}$$

[by convention, we let $\prod_{j=k}^{k-1}(I+H^jM)=1$].
Using (13) and (14), $\|E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T]\|$ can be bounded as follows:

$$\begin{aligned}
&\|E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T]\| \\
&\quad=[\text{use}(14)] \\
&\quad=\Bigg\|\prod_{j=\bar{k}}^{k-1}(I+H^jM)\cdot E[\tilde{\boldsymbol{\gamma}}^{\bar{k}}(\tilde{\boldsymbol{\gamma}}^{\bar{k}})^T] \\
&\qquad\cdot\prod_{j=\bar{k}}^{k-1}(I+H^jM)^T+\sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}(I+H^jM) \\
&\qquad\cdot H^rN(H^r)^T\cdot\prod_{j=r+1}^{k-1}(I+H^jM)^T\Bigg\| \\
&\quad\le\prod_{j=\bar{k}}^{k-1}\|I+H^jM\|^2\cdot\|E[\tilde{\boldsymbol{\gamma}}^{\bar{k}}(\tilde{\boldsymbol{\gamma}}^{\bar{k}})^T]\| \\
&\qquad+\sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}\|I+H^jM\|^2\cdot\|H^r\|^2\|N\| \\
&\quad\le[\text{use }(13)] \\
&\quad\le\prod_{j=\bar{k}}^{k-1}\left(1-\frac{\mu}{j+1}\right)\|E[\tilde{\boldsymbol{\gamma}}^{\bar{k}}(\tilde{\boldsymbol{\gamma}}^{\bar{k}})^T]\| \\
&\qquad+\sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}\left(1-\frac{\mu}{j+1}\right) \\
&\qquad\cdot\frac{1}{(r+1)^2}\|\hat{M}^{-1}\|^2\|N\|. \qquad (15)
\end{aligned}$$

To complete the proof, we shall show that the latter expression tends to 0 as $k\to\infty$. To this end, we need the following inequality:

$$\prod_{j=r+1}^{k-1}\left(1-\frac{\mu}{j+1}\right)\cdot\frac{1}{r+1}\le\frac{1}{k}\cdot\frac{(k-1)^{1-\mu}}{r^{1-\mu}} \qquad (16)$$

which we establish first. Expand the left-hand side of (16) as

$$\begin{aligned}
&\prod_{j=r+1}^{k-1}\left(1-\frac{\mu}{j+1}\right)\cdot\frac{1}{r+1} \\
&\quad=\frac{k-\mu}{k}\cdot\frac{k-1-\mu}{k-1}\cdot\frac{k-2-\mu}{k-2}\cdots\frac{r+2-\mu}{r+2}\cdot\frac{1}{r+1} \\
&\quad=\frac{1}{k}\left(1+\frac{1-\mu}{k-1}\right)\left(1+\frac{1-\mu}{k-2}\right)\cdots\left(1+\frac{1-\mu}{r+1}\right)
\end{aligned}$$

and take logarithm

$$\begin{aligned}
&\log\left(\prod_{j=r+1}^{k-1}\left(1-\frac{\mu}{j+1}\right)\cdot\frac{1}{r+1}\right) \\
&\quad=\log\frac{1}{k}+\log\left(1+\frac{1-\mu}{k-1}\right)+\cdots \\
&\qquad+\log\left(1+\frac{1-\mu}{r+1}\right) \\
&\quad\le[\text{since }\log(1+x)\le x] \\
&\quad\le\log\frac{1}{k}+\underbrace{\frac{1-\mu}{k-1}+\frac{1-\mu}{k-2}+\cdots+\frac{1-\mu}{r+1}}_{\le(1-\mu)\int_r^{k-1}\frac{1}{x}dx=(1-\mu)\log\frac{k-1}{r}=\log\frac{(k-1)^{1-\mu}}{r^{1-\mu}}} \\
&\quad\le\log\left(\frac{1}{k}\cdot\frac{(k-1)^{1-\mu}}{r^{1-\mu}}\right)
\end{aligned}$$

(16) is obtained by dropping the logarithm sign in this latter expression.

Go now back to (15) and first consider expression $\prod_{j=\bar{k}}^{k-1}(1-\frac{\mu}{j+1})$ that multiplies $\|E[\tilde{\boldsymbol{\gamma}}^{\bar{k}}(\tilde{\boldsymbol{\gamma}}^{\bar{k}})^T]\|$. We have

$$\begin{aligned}
\prod_{j=\bar{k}}^{k-1}\left(1-\frac{\mu}{j+1}\right)&=\prod_{j=\bar{k}}^{k-1}\left(1-\frac{\mu}{j+1}\right)\frac{1}{\bar{k}}\bar{k} \qquad (17)\\
&\le[\text{use (16)}] \\
&\le\frac{1}{k}\cdot\frac{(k-1)^{1-\mu}}{(\bar{k}-1)^{1-\mu}}\bar{k} \qquad (18)
\end{aligned}$$

and the last expression goes to 0 as $k\to\infty$.

Next, consider $\sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}(1-\frac{\mu}{j+1})\cdot\frac{1}{(r+1)^2}$, the expression that multiplies $\|\hat{M}^{-1}\|^2\|N\|$ in (15)

$$\begin{aligned}
&\sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}\left(1-\frac{\mu}{j+1}\right)\cdot\frac{1}{(r+1)^2} \\
&\quad=\sum_{r=\bar{k}}^{k-1}\prod_{j=r+1}^{k-1}\left(1-\frac{\mu}{j+1}\right)\cdot\frac{1}{r+1}\cdot\frac{1}{r+1} \\
&\quad\le[\text{use (16)}] \\
&\quad\le\sum_{r=\bar{k}}^{k-1}\frac{1}{k}\cdot\frac{(k-1)^{1-\mu}}{r^{1-\mu}}\cdot\frac{1}{r+1} \\
&\quad\le\frac{(k-1)^{1-\mu}}{k}\sum_{r=\bar{k}}^{k-1}\frac{1}{r^{2-\mu}} \\
&\quad\le\frac{(k-1)^{1-\mu}}{k}\int_{\bar{k}-1}^{k-1}\frac{1}{x^{2-\mu}}\,dx \\
&\quad=\frac{(k-1)^{1-\mu}}{k}\cdot\frac{1}{\mu-1}((k-1)^{\mu-1}-(\bar{k}-1)^{\mu-1})
\end{aligned}$$

Again, this expression goes to 0 as $k\to\infty$, so that both terms in the right-hand side of (15) tend to zero and this completes the proof. $*$

V. Digital Implementation

In this section, we discuss how to approximately implement the iterative identification method when the I/O are measured at sampling times only. Precisely, we suppose that the I/O data are $\{u(iT_s),\tilde{y}(iT_s)\}(i=0,1,\ldots,q)$, where T_s is sampling time satisfying $qT_s=T$.

A. Basis Functions and Reference Signal

Given functions $f_1(t),f_2(t),\ldots,f_{n+m+1}(t)$, define $V_{df}\in\mathbb{R}^{(q+1)\times(n+m+1)}$ by

$$V_{df}\triangleq\begin{bmatrix}f_1(0) & f_2(0) & \cdots & f_{n+m+1}(0)\\ f_1(T_s) & f_2(T_s) & \cdots & f_{n+m+1}(T_s)\\ \vdots & \vdots & \cdots & \vdots\\ f_1(qT_s) & f_2(qT_s) & \cdots & f_{n+m+1}(qT_s)\end{bmatrix}$$

Functions $f_i(t)$'s should be chosen such that the columns of V_{df} are linearly independent. Note also that this is true if Condition 1 is satisfied and the sampling time is fast enough.

Let the QR decomposition of V_{df} be

$$V_{df} = QR, \quad Q^TQ = I_{n+m+1}$$

where $Q \triangleq [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{n+m+1}] \in \mathbb{R}^{(q+1)\times(n+m+1)}$ and $R \in \mathbb{R}^{(n+m+1)\times(n+m+1)}$ is a nonsingular upper triangular matrix.

These $\boldsymbol{f}_i$'s constitute an orthogonal basis for projection in the digital implementation.

Next, we choose $r(t)$ and define $V_r(t), \boldsymbol{\alpha}^k$, and $\boldsymbol{\beta}^k$ by

$$V_r(t) = \left[r(t), \frac{\mathrm{d}r(t)}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}^{\max\{n,m\}}r(t)}{\mathrm{d}t^{\max\{n,m\}}}\right]$$
$$\boldsymbol{\alpha}^k \triangleq \left[1, \alpha_1^k, \ldots, \alpha_n^k\right]^T$$
$$\boldsymbol{\beta}^k \triangleq \left[\beta_0^k, \beta_1^k, \ldots, \beta_m^k\right]^T.$$

Then, $u^k(t) = A^k(p)r(t)$ and $B^k(p)r(t)$ are written as

$$A^k(p)r(t) = V_r(t)[1:n+1]\boldsymbol{\alpha}^k$$
$$B^k(p)r(t) = V_r(t)[1:m+1]\boldsymbol{\beta}^k$$

where $V_r(t)[1:r]$ is the matrix which consists of the first r columns of $V_r(t)$. Since the data are available only at sampling times, we also define

$$\boldsymbol{u}^k \triangleq [u^k(0), u^k(T_s), \ldots, u^k(qT_s)]^T \in \mathbb{R}^{q+1}$$
$$\tilde{\boldsymbol{y}}^k \triangleq [\tilde{y}^k(0), \tilde{y}^k(T_s), \ldots, \tilde{y}^k(qT_s)]^T \in \mathbb{R}^{q+1}$$

$\boldsymbol{y}^k \in \mathbb{R}^{q+1}, \boldsymbol{\varepsilon}^k \in \mathbb{R}^{q+1}$ and $\boldsymbol{w}^k \in \mathbb{R}^{q+1}$ are defined similarly. Also, we let V_{dr} be

$$V_{dr} \triangleq \begin{bmatrix} r(0) & \frac{\mathrm{d}r}{\mathrm{d}t}(0) & \cdots & \frac{\mathrm{d}^{\max\{n,m\}}r}{\mathrm{d}t^{\max\{n,m\}}}(0) \\ r(T_s) & \frac{\mathrm{d}r}{\mathrm{d}t}(T_s) & \cdots & \frac{\mathrm{d}^{\max\{n,m\}}r}{\mathrm{d}t^{\max\{n,m\}}}(T_s) \\ \vdots & \vdots & \cdots & \vdots \\ r(qT_s) & \frac{\mathrm{d}r}{\mathrm{d}t}(qT_s) & \cdots & \frac{\mathrm{d}^{\max\{n,m\}}r}{\mathrm{d}t^{\max\{n,m\}}}(qT_s) \end{bmatrix}$$

Then, we have

$$\boldsymbol{u}^k = V_{dr}[1:n+1]\boldsymbol{\alpha}^k.$$

B. Computation of $\boldsymbol{\delta}^k$

Since $\boldsymbol{f}_i \ \ (i = 1, 2, \ldots, n+m+1)$ is an orthogonal set of vectors with unitary norm, the projection of $\boldsymbol{\varepsilon}^k$ onto $\mathrm{span}\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{n+m+1}\}$ is $[\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{n+m+1}]Q^T\boldsymbol{\varepsilon}^k = V_{df}R^{-1}Q^T\boldsymbol{\varepsilon}^k$, where $R^{-1}Q^T\boldsymbol{\varepsilon}^k$ is an approximate expression for $\boldsymbol{\delta}^k$ in (2). This suggests using

$$R^{-1}Q^T\boldsymbol{\varepsilon}^k$$

in place of $\boldsymbol{\delta}^k$ in (3).

C. Estimation of M and Overall Implementation

Note that, if sampling time T_s is fast enough so that $u(t)$ does not change significantly between samples, we can approximately write

$$\boldsymbol{y}^k \simeq G\boldsymbol{u}^k \tag{19}$$

where $G \in \mathbb{R}^{(q+1)\times(q+1)}$ is a system-dependent Toeplitz matrix of the form

$$G = \begin{bmatrix} g_0 & 0 & 0 & \ldots & 0 \\ g_1 & g_0 & 0 & \ldots & 0 \\ g_2 & g_1 & g_0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ g_q & g_{q-1} & g_{q-2} & \cdots & g_0 \end{bmatrix}$$

Thus, the mismatch signal $\boldsymbol{\varepsilon}^k$ can also be expressed as

$$\boldsymbol{\varepsilon}^k \simeq GV_{dr}[1:n+1]\boldsymbol{\alpha}^k - V_{dr}[1:m+1]\boldsymbol{\beta}^k + \boldsymbol{w}^k. \tag{20}$$

Substituting this expression for $\boldsymbol{\varepsilon}^k$ in $\boldsymbol{\delta}^k \simeq R^{-1}Q^T\boldsymbol{\varepsilon}^k$ yields $\boldsymbol{\delta}^k \simeq R^{-1}Q^TGV_{dr}[1:n+1]\boldsymbol{\alpha}^k - R^{-1}Q^TV_{dr}[1:m+1]\boldsymbol{\beta}^k + R^{-1}Q^T\boldsymbol{w}^k$. Comparing with (4), we then obtain

$$M \simeq R^{-1}Q^T[GV_{dr}[2:n+1], -V_{dr}[1:m+1]] \tag{21}$$
$$\bar{\boldsymbol{\delta}} \simeq R^{-1}Q^TGV_{dr}[1]. \tag{22}$$

Also, recall that $\bar{\boldsymbol{\delta}} = -M\boldsymbol{\gamma}^\circ$, from which

$$\boldsymbol{\gamma}^\circ = -M^{-1}\bar{\boldsymbol{\delta}}. \tag{23}$$

The equations we have derived suggest the following practical implementation of the identification algorithm. Data are first used to obtain a rough estimate of the g_i's, that is, the Markov coefficients of system (19), and, through this, a rough estimate $\hat{G}$ of G is obtained. Then, estimates $\hat{M}$ of M and $\hat{\bar{\boldsymbol{\delta}}}$ of $\bar{\boldsymbol{\delta}}$ are derived from (21) and (22) by substituting G with $\hat{G}$. $\hat{M}$ is used in the gain of the ILC procedure. Moreover, a rough initial estimate $\boldsymbol{\gamma}^0 = -\hat{M}^{-1}\bar{\boldsymbol{\delta}}$ for $\boldsymbol{\gamma}^\circ$ is obtained [see (23)]. With $\hat{M}$ and $\boldsymbol{\gamma}^0$ in our hands, the iterative ILC procedure is then used to refine the estimate $\boldsymbol{\gamma}^k$.

Remark 4: Alternative indirect methods can be used in the identification of continuous-time systems where one first converts the model into a discrete-time version of it, then performs identification with a standard discrete-time algorithm, and finally converts the system back to continuous-time.

The chief recognized drawback of this indirect approach is that the final result becomes more and more sensitive to noise as sampling is made faster; see, e.g., [12] for a thorough discussion on the dependability of indirect methods. Similarly, direct methods where time derivatives of I/O data are used suffer from the same drawback. The ILC approach presented herein overcomes this drawback because no derivatives of measured signals are considered. The fact that this drawback is overcome can also be directly argued from an inspection of (20): this equation shows that noise enters the computation of $\boldsymbol{\varepsilon}^k$—which is the driving term in the updating equation for parameter estimation—in a way that is unaffected by the sampling time T_s.

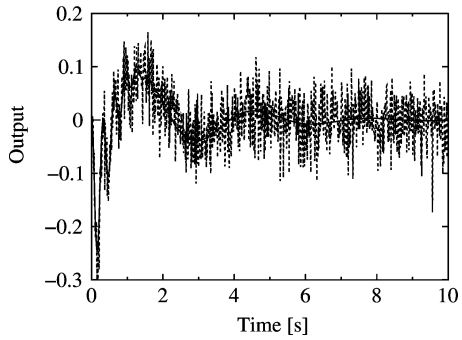

Fig. 2. Output with noise for impulse-like input.

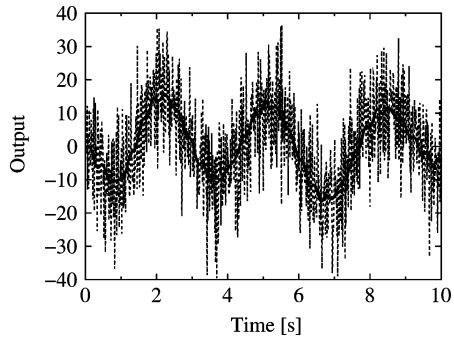

Fig. 3. Measured output $\bar{y}(t)$ and $B^k(p)r(t)$ at 30th trial (NSR: 100%).

As an additional remark, we note that indirect methods require regular sampling whereas the digital implementation of the ILC approach discussed in this section can be easily generalized to the case of unequally spaced sampling. $*$

VI. Numerical Example

The effectiveness of the proposed method will be evaluated through simulation in this section.

Consider a linear, fourth-order, nonminimum phase system with complex poles described by the transfer function (this is the Rao–Garnier test system [12])

$$P(s) = \frac{-T_1 s + 1}{\left(\frac{s^2}{\omega_{n,1}^2} + \frac{2\zeta_1 s}{\omega_{n,1}} + 1\right)\left(\frac{s^2}{\omega_{n,2}^2} + \frac{2\zeta_2 s}{\omega_{n,2}} + 1\right)}$$

where $T_1 = 4$ [s], $\omega_{n,1} = 20$ [rad/s], $\zeta_1 = 0.1$, $\omega_{n,2} = 2$ [rad/s], and $\zeta_2 = 0.25$. The time span for each trial is $T = 10$ [s], and the sampling time is $T_s = 10$ [ms]. Namely, the number of data used for one trial is 1001.

With the goal of estimating the g_i's, an impulse-like input was injected in the system. Precisely, we let $u(t) = 1$ in the interval 0–10 [ms], and then the input was set to zero. Fig. 2 shows the measured output [noise-to-signal ratio (NSR) was 100%, where NSR is defined as $\mathrm{NSR} \triangleq \|\boldsymbol{w}\|/\|\boldsymbol{y}\|$]. We took $\hat{g}_i = \tilde{y}(iT_s)(i = 0, 1, \ldots, q)$ and estimated M and $\boldsymbol{\gamma}^\circ$ as indicated at the end of the previous section.

We next proceeded to iterative identification. $r(t)$ was obtained as the output of the system

$$F(s) = \frac{10^6}{(s+10)^6}$$

when the input is chosen as the following sum of five sinusoidal signals:

$$\sin(t) + \sin(1.9t) + \sin(2.1t) + \sin(18t) + \sin(22t) \quad (24)$$

and the system $F(s)$ has zero initial conditions, so that signal $r(t)$ has zero initial derivatives. We also let $f_1(t) = r(t), f_2(t) = \mathrm{d}r(t)/\mathrm{d}t, \ldots, f_{n+m+1}(t) = \mathrm{d}^{n+m+1}r(t)/\mathrm{d}t^{n+m+1}$. The measurement noise was white with zero mean and variance σ^2. The variance σ^2 was chosen so that the NSR was 100%, where NSR is defined as $\mathrm{NSR} \triangleq \|\boldsymbol{w}\|/\|\boldsymbol{y}^{30}\|$.

We used the update law (3) where H^k is given by (8) and (9) with M estimated as indicated before and where $10I$ was used for N, which is a very coarse estimate of the true N which was about ten times larger. Fig. 3 shows the output $\tilde{y}^k(t)$ obtained at the 30th trial. The same figure also displays the signal $B^k(p)r(t)$ (thick line). As it can be seen, $y^k(t)$ tracks $B^k(p)r(t)$ very well despite the heavy measurement noise, which indicates that the plant model is identified accurately. In fact, an accurate description of the plant is obtained after few iterations. To confirm this, Bode plots of the estimated system at the 3rd trial are shown in Fig. 4 for 50 runs of the algorithm (bundle of thin lines) against the Bode plot of the true system. For comparison, the results achieved with the SRIVC method in the CONTSID tool-box are shown in Fig. 5 when the input is (24), $T = 100$ [s] and T_s is as before.

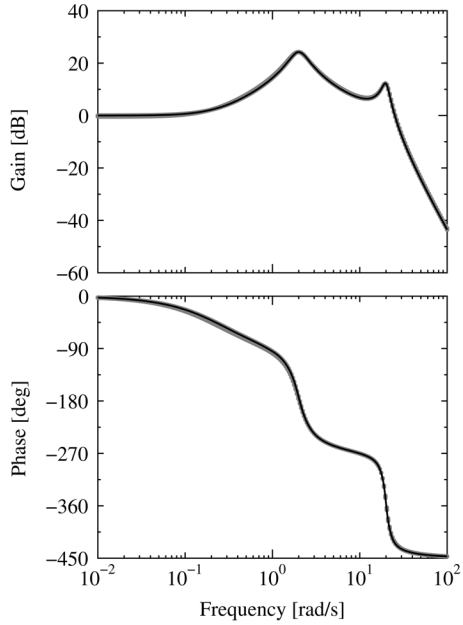

Fig. 4. Bode plots of the estimated system $(k = 3)$.

Fig. 6 shows an example of the estimated coefficients through trials. From this figure, we see that the initial estimation is improved quickly. The Euclidean norm of the parameter estimation error is shown in Fig. 7.

VII. Conclusion

In this paper, a novel approach has been introduced for the identification of linear continuous-time systems based on the iterative learning control concept. The method achieves identification through repetition of trials in the presence of heavy

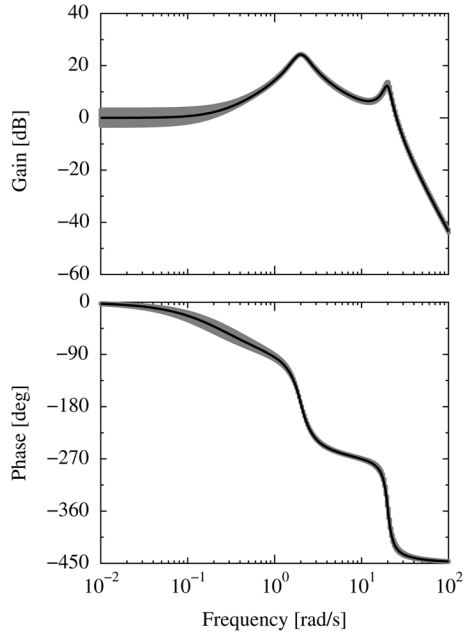

Fig. 5. Bode plots of the system estimated with SRIVC.

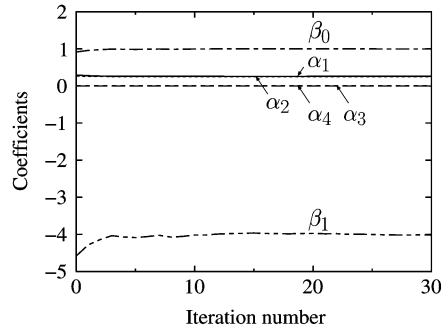

Fig. 6. Identified coefficients γ^k in each trial (True value: $\alpha_1 = 0.26, \alpha_2 = 0.255, \alpha_3 = 0.003125, \alpha_4 = 0.000625, \beta_0 = 1, \beta_1 = -4$).

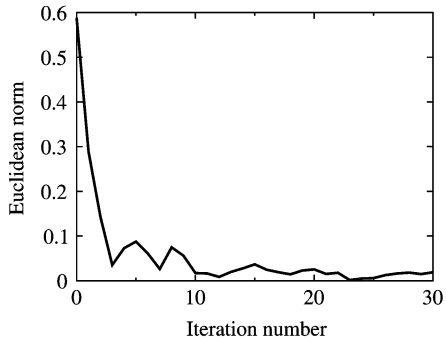

Fig. 7. Euclidean norm of the parameter estimation error $\|\gamma^k - \gamma^\circ\|$.

measurement noise. The proposed method has several advantages among which that no time-derivatives of I/O data is required and no data preprocessing (e.g., decimation or filtering) is necessary. The effectiveness of the proposed method has been demonstrated through numerical examples for a nonminimum phase plant.

REFERENCES

[1] S. Arimoto, *Control Theory of Non-Linear Mechanical Systems: A Passivity-Based and Circuit-Theoretic Approach*. Oxford, U.K.: Oxford Univ. Press, 1996, ch. 4 and 5.
[2] S. Arimoto, S. Kawamura, and F. Miyazaki, "Bettering operation of robotics.," *J. Robot. Syst.*, vol. 1, no. 2, pp. 123–140, 1984.
[3] Z. Bein and J. X. Xu, *Iterative Learning Control—Analysis, Design, Integration and Applications*. Norwell, MA: Kluwer, 1998.
[4] Y. Chen and C. Wen, *Iterative Learning Control: Convergence, Robustness and Applications*. New York: Springer-Verlag, 1999, vol. LNCIS-248.
[5] H. Garnier, M. Gilson, and E. Husestein, "Developments for the Matlab CONTSID toolbox.," in *Proc. IFAC Symp. System Identification*, Rotterdam, The Netherlands, 2003.
[6] H. Garnier and M. Mensler, "CONTSID: A continuous-time system identification toolbox for Matlab.," in *Proc. 5th Eur. Control Conf.*, Karlsruhe, Germany, 1999.
[7] H. Garnier and M. Mensler, "The CONTSID toolbox: A Matlab toolbox for continuous-time system identification.," in *Proc. 12th IFAC Symp. System Identification*, Santa Barbara, CA, 2000.
[8] K. Hamamoto and T. Sugie, "An iterative learning control algorithm within prescribed input-output subspace.," *Automatica*, vol. 37, no. 11, pp. 1803–1809, 2001.
[9] K. Hamamoto and T. Sugie, "Iterative learning control for robot manipulators using the finite dimensional input subspace.," *IEEE Trans. Robot. Autom,*, vol. 18, no. 4, pp. 632–635, Aug. 2002.
[10] S. Kawamura, F. Miyazaki, and S. Arimoto, "Realization of robot motion based on a learning method.," *IEEE Trans. Syst., Man., Cybern.*, vol. 18, no. 1, pp. 126–134, Jan./Feb. 1988.
[11] K. L. Moore, *Iterative Learning Control for Deterministic Systems.*, ser. Advances in Industrial Control. London, U.K.: Springer-Verlag, 1993.
[12] G. P. Rao and H. Garnier, "Identification of continuous-time systems: Direct or indirect?," *Syst. Sci.*, vol. 30, no. 3, pp. 25–50, 2004.
[13] F. Sakai and T. Sugie, "Continuous-time systems identification based on iterative learning control.," in *Proc. 16th IFAC World Congr.*, Prague, Czech Republic, 2005.
[14] , N. K. Sinha and G. P. Rao, Eds., *Identification of Continuous-Time Systems. Methodology and Computer Implementation.*. Dordrecht, The Netherlands: Kluwer, 1991.
[15] T. Sugie and T. Ono, "An iterative learning control law for dynamical systems.," *Automatica*, vol. 27, no. 4, pp. 729–732, 1991.
[16] T. Sugie and F. Sakai, "Noise tolerant iterative learning control for continuous-time systems identification.," in *Proc. 44th IEEE Conf. Decision and Control and Eur. Control Conf.*, Seville, Spain, 2005.
[17] H. Unbehauen and G. P. Rao, "Continuous-time approaches to system identification—A survey.," *Automatica*, vol. 26, no. 1, pp. 23–35, 1990.
[18] J. X. Xu, "The frontiers of iterative learning control II," *Syst., Control, Inf.*, vol. 46, no. 2, pp. 233–243, 2002.
[19] P. Young, "Parameter estimation for continuous-time models—A survey.," *Automatica*, vol. 17, no. 1, pp. 23–39, 1981.

Marco C. Campi (SM'08) was born in Tradate, Italy, on December 7, 1963. In 1988, he received the Ph.D. degree in electronic engineering from the Politecnico di Milano, Milano, Italy.

He is a Professor of automatic control at the University of Brescia, Brescia, Italy, where he has been since 1992. From 1988 to 1989, he was a Research Assistant at the Department of Electrical Engineering of the Politecnico di Milano. From 1989 to 1992, he worked as a Researcher at the Centro di Teoria dei Sistemi of the National Research Council (CNR) in Milano. He has held visiting and teaching positions at many universities and institutions including the Australian National University, Canberra, Australia; the University of Illinois at Urbana-Champaign; the Centre for Artificial Intelligence and Robotics, Bangalore, India; the University of Melbourne, Melbourne, Australia; and Kyoto University, Kyoto, Japan. He is an Associate Editor of *Systems and Control Letters* and a past Associate Editor of *Automatica* and the *European Journal of Control*. His research interests include: system identification, stochastic systems, adaptive and data-based control, robust convex optimization, robust control and estimation, and learning theory.

Dr. Campi received the "Giorgio Quazza" prize as the Best Original Doctoral Thesis for 1988. He serves as Chair of the Technical Committee IFAC on Stochastic Systems (SS) and is a member of the Technical Committee IFAC on Modeling, Identification, and Signal Processing (MISP) and of the Technical Committee IFAC on Cost Oriented Automation. Moreover, he is a Distinguished Lecturer under the IEEE Control Systems Society (CSS) Program.

Toshiharu Sugie (M'89–SM'02–F'07) received the B.E., M.E., and Ph.D. degrees in engineering form Kyoto University, Kyoto, Japan, in 1976, 1978, and 1985, respectively.

From 1978 to 1980, he was a Research Member of Musashino Electric Communication Laboratory in NTT, Musashino, Japan. From 1984 to 1988, he was a Research Associate of the Department of Mechanical Engineering, University of Osaka Prefecture, Osaka, Japan. In 1988, he joined Kyoto University, where he is currently a Professor in the Department of Systems Science. His research interests are in robust control, learning control, nonlinear control, identification for control, and control application to mechanical systems.

Dr. Sugie is a Fellow of the Society of Instrument and Control Engineers, Japan. He is an Editor of *Automatica* and was also an Associate Editor of the *Asian Journal of Control* and the *International Journal of Systems Science*.

Fumitoshi Sakai (M'05) was born in Tochigi, Japan, in 1973. He received the B.S., M.S., and Ph.D. degrees in engineering from Kanazawa University, Kanazawa, Japan, in 1996, 1998, and 2001, respectively.

In 2001, he joined the Department of Mechanical Engineering, Nara National College of Technology, Nara, Japan, as a Research Associate. His research interests include system identification and learning control.