Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006

Iterative identification method for linear continuous-time systems

Marco C. Campi, Toshiharu Sugie and Fumitoshi Sakai

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

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 be convenient to have continuous-time models directly from the sampled I/O data. Indeed most design tools are suitable for continuous-time systems and it is easy for us to capture the plant dynamics intuitively in continuous-time rather than in discrete-time.

The basic difficulty of the direct identification approach lies in handling the time-derivatives of I/O data in the presence of measurement noise. A lot of effort has been made to circumvent the need to reconstruct these time derivatives. A comprehensive survey of these techniques has been first given by [19] and then by [17]. A book has also been devoted to these so-called direct methods [14]. Furthermore the Continuous-Time System Identification (CONTSID) toolbox has been developed on the basis of these methods [6], [7], [5].

On the other hand, iterative learning control (ILC) has attracted much attention over the last two decades as a powerful model-free control method [2], [12], [10], [3], [4], [18], [1]. ILC returns the desired input which achieves perfect 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], have proposed an ILC where learning law works in the prescribed finite-dimensional subspace, showing that any time-derivative of tracking error is not required to achieve perfect tracking for uncertain systems in continuous-time. Based on this work, Sugie and Sakai [13], [16] proposed an ILC which works in the presence of heavy measurement noise (such as more than 30% noise to signal ratio), and the method is shown to be applicable to identification of continuous-time systems. This method has several advantages such as (i) no time-derivatives of I/O data is required, (ii) unbiased estimation is obtained, and (iii) the identified model quality can be estimated by watching the tracking control performance in experiments. The system class, however, is restricted to those having no zeros (*i.e.,* all pole types). Therefore, from the view point of identification, this restriction should be removed.

The purpose of this paper is to propose an identification method for linear continuous-time systems having poles and zeros based on the ILC approach proposed in [13], [16]. Moreover, differently from [13] and [16], the method guarantees zero convergence of the parameter estimation errors as the trial number increases.

The following notation will be used. The superscript of the variables denotes the trial number and the subscript denotes the element number of a set or a matrix. Namely, the input u of the kth trial is denoted by u^k and the ith element of the vector x is denoted by x_i.

II. SYSTEM DESCRIPTION

Consider the continuous-time SISO system described by

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

where $u(t) \in L_2[0,\ T]$ and $y(t) \in L_2[0,\ T]$ are the input and the output, respectively. $\alpha_i^\circ \in \mathbb{R}\ (i = 0, 1, \cdots, n)$ and $\beta_i^\circ \in \mathbb{R}\ (i = 0, 1, \cdots, m)$ are coefficient parameters, while p is the differential operator, *i.e.,* $pu(t) = \mathrm{d}u(t)/\mathrm{d}t$. We assume the following:

- The system has 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.
- We can measure $\tilde{y}(t)$, the output contaminated with noise,
$$\tilde{y}(t) = y(t) + w(t) \quad (2)$$
and $w(t)$ is zero-mean measurement noise.
- We can repeat the experiments at the same initial condition on the time interval $[0,\ T]$.

The goal is to find the system model among the model class described by

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

Marco Campi is with Dipartimento di Elettronica per l'Automazione, Università di Brescia, via Branze 38, 25123 Brescia, Italy campi@ing.unibs.it

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

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

based on the sampled I/O data $\{u(iT_s), \tilde{y}(iT_s)\}$ $(i = 0, 1, 2, \cdots, N)$, where T_s denotes the sampling period satisfying $NT_s = T$.

III. Identification procedure

This section gives an identification procedure through iteration of trials.

A. Data generation scheme and parameter update law

First, choose a signal $r(t) \in L_2[0, T]$ which is at least n times continuously differentiable. Then, perform the following experiment at the k-th trial as shown in Fig. 1, which produces the signal $\varepsilon^k(t)$ when the parameter estimates $\alpha_1^k, \cdots, \alpha_n^k, \beta_0^k, \cdots, \beta_m^k$ from the previous trial are given.

(i) Define

$$\begin{aligned} A^k(p) &= 1 + \alpha_1^k p + \cdots + \alpha_n^k p^n \\ B^k(p) &= \beta_0^k + \beta_1^k p + \cdots + \beta_m^k p^m \end{aligned}$$

(ii) Generate $u^k(t) = A^k(p)r(t)$.
(iii) Inject $u^k(t)$ into the system, and collect $\tilde{y}^k(t)$.
(iv) Generate $B^k(p)r(t)$.
(v) Obtain the mismatch signal $\varepsilon^k(t)$ by

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

Note that $\varepsilon^k(t)$ is obtained without taking any derivative of noisy measurements, only derivatives of $r(t)$ are required.

Now, we introduce $n + m + 1$ basis functions $f_1(t), \ldots, f_{n+m+1}(t) \in L_2[0, T]$ which satisfy the following condition.

Condition 1: For $B(p)/A(p) \in \mathcal{M}$, if

$$\int_0^T \left(\frac{B^\circ(p)A(p) - B(p)A^\circ(p)}{A^\circ(p)} r(t) \right) \cdot f_i(t) = 0, \quad i = 1, \ldots, n+m+1, \tag{3}$$

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

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.

Then, project $\varepsilon^k(t)$ onto the finite-dimensional subspace described by

$$\mathcal{F} \triangleq \text{span}\{f_1(t), f_2(t), \cdots, f_{n+m+1}(t)\}$$

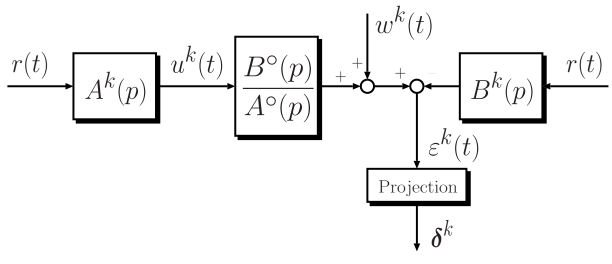

Fig. 1. Data generation scheme at k-th tried

Let the projection of $\varepsilon^k(t)$ onto $\mathcal{F}$ be

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

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

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

$$\begin{aligned} \boldsymbol{\gamma}^\circ &= [\alpha_1^\circ, \cdots, \alpha_n^\circ, \beta_0^\circ, \cdots, \beta_m^\circ]^T, \\ \boldsymbol{\gamma}^k &= [\alpha_1^k, \cdots, \alpha_n^k, \beta_0^k, \cdots, \beta_m^k]^T, \end{aligned}$$

then the iterative identification procedure is described as follows where $\epsilon_0 > 0$ is chosen by the designer:

Step 0) Given $\boldsymbol{\gamma}^0$, set $k = 0$.
Step 1) Generate $\boldsymbol{\delta}^k$ from $\boldsymbol{\gamma}^k$ according to the scheme as shown in Fig. 1.
Step 2) Update $\boldsymbol{\gamma}^k$ by the following rule.

$$\boldsymbol{\gamma}^{k+1} = \boldsymbol{\gamma}^k + H^k \boldsymbol{\delta}^k \tag{4}$$

where H^k is the learning gain. If $\|\boldsymbol{\gamma}^{k+1} - \boldsymbol{\gamma}^k\| \leq \epsilon_0$, stop the iteration. 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 identification procedure defines an affine operator from $\mathbb{R}^{n+m+1}$ (space for $\boldsymbol{\gamma}^k$) and $\mathbb{R}^{n+m+1}$ (space for $\boldsymbol{\delta}^k$), that is

$$\boldsymbol{\delta}^k = M\boldsymbol{\gamma}^k + \bar{\boldsymbol{\delta}}^k, \tag{5}$$

where M is a $(n+m+1) \times (n+m+1)$ matrix and $\bar{\boldsymbol{\delta}}^k$ is the offset term. Note, from (3) we know that $\boldsymbol{\delta}^k = 0$ implies $\boldsymbol{\gamma}^k = \boldsymbol{\gamma}^\circ$. This means that the equation

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

has the only solution $\boldsymbol{\gamma}^k = \boldsymbol{\gamma}^\circ$, so that M is non-singular and $\bar{\boldsymbol{\delta}}^k = -M\boldsymbol{\gamma}^\circ$. Thus, (5) can be re-written as

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

with M non-singular. When noise $w^k(t)$ is taken into account, $\boldsymbol{\delta}^k$ becomes

$$\boldsymbol{\delta}^k = M(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ) + \boldsymbol{\nu}^k, \tag{6}$$

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

Eqns. (4) and (6) yield

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

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

$$\tilde{\boldsymbol{\gamma}}^{k+1} = (I + H^k M)\tilde{\boldsymbol{\gamma}}^k + H^k \boldsymbol{\nu}^k \tag{7}$$

which is the equation that describes how the error $\boldsymbol{\delta}^k$ 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 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. Noise is assumed to be independent in different experiments.

From (7), we have:

$$\begin{aligned} P^{k+1} &= E\left[\{(I+H^kM)\tilde{\boldsymbol{\gamma}}^k + H^k\boldsymbol{\nu}^k\}\right.\\ &\qquad\qquad \left.\{(I+H^kM)\tilde{\boldsymbol{\gamma}}^k + H^k\boldsymbol{\nu}^k\}^T\right]\\ &= (I+H^kM)P^k(I+H^kM)^T + H^kN(H^k)^T. \quad (8)\end{aligned}$$

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

$$H^k = -P^kM^T(MP^kM^T+N)^{-1}. \quad (9)$$

This latter equation should be initialized with $P^0 = E[\tilde{\boldsymbol{\gamma}}^0(\tilde{\boldsymbol{\gamma}}^0)^T]$. If such a P^0 is not known and a conventional $P^0 > 0$ is instead used, optimality is lost but the convergence result in Theorem 1 below can be shown to be still valid.

With this choice, we obtain

$$P^{k+1} = P^k - P^kM^T(MP^kM^T+N)^{-1}MP^k. \quad (10)$$

Eqns. (9) and (10) give the way to select H^k.

For this choice of H^k, we obtain the following result.

Theorem 1: If the updating law (4) with (9) and (10) is adopted, then

$$P^k = E[(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)(\boldsymbol{\gamma}^k - \boldsymbol{\gamma}^\circ)^T] \to 0, \quad \text{as} \quad k\to\infty. \quad (11)$$

Proof:

In (8) take $H^k = -\rho M^{-1}$ with $0<\rho<1$ (instead of (9)) and denote the corresponding solution by P^k_ρ. Let $P^0_\rho = P^0$. Then we have

$$P^{k+1}_\rho = (1-\rho)^2P^k_\rho + M^{-1}N(M^{-1})^T\rho^2,$$

from which

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

holds and the first term vanishes as k grows.

Now we claim that $P^k \le P^k_\rho$ holds for any k. By induction $P^0 = P^0_\rho$. Assume $P^k \le P^k_\rho$. Then we obtain

$$\begin{aligned} P^{k+1} &= (I+H^kM)P^k(I+H^kM)^T + H^kN(H^k)^T\\ &\le \left[\text{since } H^k \text{ in (9) is optimal}\right]\\ &\le (I-\rho M^{-1}M)P^k(I-\rho M^{-1}M)^T\\ &\qquad\qquad + \rho M^{-1}N\rho(M^{-1})^T\\ &= (1-\rho)^2P^k + M^{-1}N(M^{-1})^T\rho^2\\ &\le (1-\rho)^2P^k_\rho + M^{-1}N(M^{-1})^T\rho^2\\ &= P^{k+1}_\rho,\end{aligned}$$

so closing the induction. Consequently, we have

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

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. Thus, the left-hand-side has to be zero. (Q.E.D.)

This theorem implies that the proposed method gives us the true parameters $\boldsymbol{\gamma}^\circ$ in the presence of measurement noise through iteration of trials.

IV. Simplified learning gain

The gain determined by (9) and (10) 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. It performs an optimal compromise between exploitation of information and rejection of noise. In the case when $N=0$ (no noise) we have:

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

so that

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

and error goes to zero in one step.

Simpler updating rules for H^k can be used, however, that still guarantee the error convergence to zero. One such rule is given by

$$H^k = -\frac{1}{k+1}M^{-1} \quad (12)$$

A. Convergence analysis

Theorem 2: If we adopt the updating law (4) with (12), then

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

Proof: Since

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

holds, we have

$$\begin{aligned} &E[\tilde{\boldsymbol{\gamma}}^{k+1}(\tilde{\boldsymbol{\gamma}}^{k+1})^T]\\ &= \frac{k^2}{(k+1)^2}E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T] + \frac{1}{(k+1)^2}M^{-1}N(M^{-1})^T\\ &= \frac{1}{(k+1)^2}M^{-1}N(M^{-1})^T\\ &\qquad + \frac{k^2}{(k+1)^2}\cdot\frac{1}{k^2}M^{-1}N(M^{-1})^T\\ &\qquad + \cdots\\ &\qquad + \frac{k^2}{(k+1)^2}\cdot\frac{(k-1)^2}{k^2}\cdots\frac{1}{1^2}M^{-1}N(M^{-1})^T\\ &= \underbrace{\left[\frac{1}{(k+1)^2} + \cdots + \frac{1}{(k+1)^2}\right]}_{(k+1)-\text{times}}M^{-1}N(M^{-1})^T\\ &= \frac{1}{k+1}M^{-1}N(M^{-1})^T. \quad (13)\end{aligned}$$

The above term vanishes as $k \to \infty$, which proves the theorem. (Q.E.D.)

Remark 2: One important merit of (12) is stated as follows. From (13), it is obvious that if $N = 0$ (no noise) we get $E[\tilde{\boldsymbol{\gamma}}^1(\tilde{\boldsymbol{\gamma}}^1)^T] = 0$, *i.e.,* error goes to zero in one step like for the optimal choice (9) and (10). The reason why this happens is that $H^1 = -M^{-1}$ so that $\tilde{\boldsymbol{\gamma}}^1 = (I - M^{-1}M)\boldsymbol{\gamma}^0 - M^{-1}\boldsymbol{\nu}^0 = -M^{-1}\boldsymbol{\nu}^0$ and initial error is killed in 1 step. This is not immediate in other contexts where the observation matrix (M here) is not invertible. Note also that, $E[\tilde{\boldsymbol{\gamma}}^{k+1}(\tilde{\boldsymbol{\gamma}}^{k+1})^T]$ depends linearly on N. This is important since we expect that N is small because $\boldsymbol{\nu}^k$ is obtained via projection.

B. Robustness analysis

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

$$H^k = -\frac{1}{k+1}\hat{M}^{-1} \tag{14}$$

in place of (12). We next want to investigate conditions under which $E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T] \to 0$ still holds using (14).

Theorem 3: If $x^T\hat{M}^{-1}Mx > 0$ holds for $\forall x \neq 0$, then $E[\tilde{\boldsymbol{\gamma}}^k(\tilde{\boldsymbol{\gamma}}^k)^T] \to 0$ with the choice of (14).

Remark 3: In the scalar case $x^T\hat{M}^{-1}Mx > 0$ suit means that M and $\hat{M}$ have the the same sign.

Proof: Omitted due to space limitations. (Q.E.D.)

V. Digital implementation

In this section, we discuss how to implement the iterative identification method when the I/O data are available only on the sampled time and the input is injected to the plant via a zero-order holder.

A. Basis and reference signal

First, we take a "rich" signal $s_0(t) \in L_2[0,T]$, which is $n+m$ times continuously differentiable, and define

$$\begin{aligned} V_s(t) &= \left[s_0(t), \frac{\mathrm{d}s_0(t)}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}^{n+m}s_0(t)}{\mathrm{d}t^{n+m}}\right] \\ &\triangleq [s_0(t), s_1(t), \ldots, s_{n+m}(t)]. \end{aligned}$$

In addition, for digital implementation, we define $V_{ds} \in \mathbb{R}^{(N+1)\times(n+m+1)}$ by

$$V_{ds} \triangleq \begin{bmatrix} s_0(0) & s_1(0) & \ldots & s_{n+m}(0) \\ s_0(T_s) & s_1(T_s) & \ldots & s_{n+m}(T_s) \\ \vdots & \vdots & \ldots & \vdots \\ s_0(NT_s) & s_1(NT_s) & \ldots & s_{n+m}(NT_s) \end{bmatrix}$$

and let the QR decomposition of V_{ds} be

$$V_{ds} = UR, \quad U^TU = I_{n+m+1}, \tag{15}$$

where $U \triangleq [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{n+m+1}] \in \mathbb{R}^{(N+1)\times(n_f+1)}$ and $R \in \mathbb{R}^{(n+m+1)\times(n+m+1)}$ is a nonsingular upper triangular matrix. The signal $s_0(t)$ is supposed to satisfy

$$\mathrm{rank}(V_{ds}^TV_{ds}) = n+m+1. \tag{16}$$

These $\boldsymbol{f}_i$'s constitute the orthogonal basis of the projection.

Next, we choose $r(t)$ which is at least n times continuously differentiable, and define $V_r(t)$, $\boldsymbol{\alpha}^k$ and $\boldsymbol{\beta}^k$ by

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

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

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

where $V_r(t)[1:m+1]$ means the vector which consists of the columns of 1 through $(m+1)$ of $V_r(t)$. Since the data are available only at sampled times, we define

$$\begin{aligned} \boldsymbol{u}^k &\triangleq [u^k(0), u^k(T_s), \cdots, u^k(NT_s)]^T \in \mathbb{R}^{N+1}, \\ \tilde{\boldsymbol{y}}^k &\triangleq [\tilde{y}^k(0), \tilde{y}^k(T_s), \cdots, \tilde{y}^k(NT_s)]^T \in \mathbb{R}^{N+1}. \end{aligned}$$

$\boldsymbol{y}^k \in \mathbb{R}^{N+1}$, $\boldsymbol{\varepsilon}^k \in \mathbb{R}^{N+1}$ and $\boldsymbol{w}^k \in \mathbb{R}^{N+1}$ are defined in the same way. Similarly, we let $V_{dr} \in \mathbb{R}^{(N+1)\times(n+m+1)}$ be

$$V_{dr} \triangleq \begin{bmatrix} r(0) & \dot{r}(0) & \ldots & r^{(n)}(0) \\ r(T_s) & \dot{r}(T_s) & \ldots & r^{(n)}(T_s) \\ \vdots & \vdots & \ldots & \vdots \\ r(NT_s) & \dot{r}(NT_s) & \ldots & r^{(n)}(NT_s) \end{bmatrix}.$$

Then, we have

$$\boldsymbol{u}^k = V_{dr}\boldsymbol{\alpha}^k.$$

B. Estimate of M

In this subsection, we will show how to estimate M from the available I/O data.

Since, the plant (1) is SISO LTI, $\boldsymbol{y}^k$ and $\boldsymbol{u}^k$ are linked by

$$\boldsymbol{y}^k = G\boldsymbol{u}^k \tag{17}$$

irrespective of k, where $G \in \mathbb{R}^{(N+1\times N+1)}$ is a system-dependant matrix of the form

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

The first column of G is the output $\boldsymbol{y}$ when we inject the input $\boldsymbol{u} = [1, 0, 0, \cdots, 0]^T$ held with a zero-order holder.

The mismatch signal $\boldsymbol{\varepsilon}^k$ is obtained as

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

Since $\boldsymbol{f}_i$ $(i = 1, 2, \ldots, n+m+1)$ is an orthogonal basis, the projection of $\boldsymbol{\varepsilon}^k$ onto $\mathrm{span}\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{n+m+1}\}$ is

$$\boldsymbol{\delta}^k = U^T\boldsymbol{\varepsilon}^k.$$

Therefore, M and $\bar{\boldsymbol{\delta}}^k$ in (5) are given by

$$M = U^T\left[GV_{dr}[2:n+1], \ -V_{dr}[1:m+1]\right], \tag{18}$$

$$\bar{\boldsymbol{\delta}}^k = U^TGV_{dr}[1]. \tag{19}$$

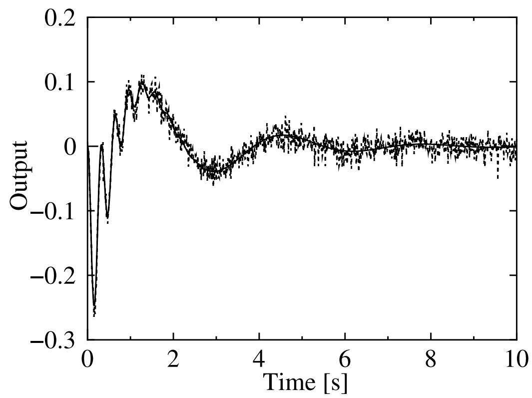

Fig. 2. Output with noise for impulse-like input

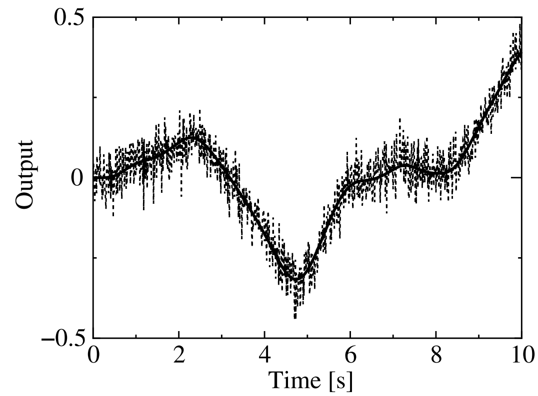

Fig. 3. Measured output $\tilde{y}(t)$ and $B^k(p)r(t)$ at 5th trial (NSR:30%)

Also, the initial estimate of the parameter vector, $\boldsymbol{\gamma}^0$, is given by

$$\boldsymbol{\gamma}^0 = -M^{-1}\bar{\boldsymbol{\delta}}^k. \tag{20}$$

Clearly, eqns. (18) and (20) cannot be directly used to compute M and $\boldsymbol{\gamma}^0$ since they contain G which in turn depends on the unknown system. One way to obtain an estimate of G, say $\hat{G}$, is to get the data g_i $(i = 1, \cdots, N)$ through an experiment.

VI. Numerical example

We will evaluate the effectiveness of the proposed method through simulation in this subsection.

Consider a linear, fourth-order, non-minimum phase system with complex poles whose transfer function (The Rao-Garnier test system [11]) is given by:

$$P(s) = \frac{K(-T_1 s + 1)}{\left(\frac{s^2}{\omega_{n,1}^2} + \frac{2\zeta_1 s}{\omega_{n,1}} + 1\right)\left(\frac{s^2}{\omega_{n,2}^2} + \frac{2\zeta_2 s}{\omega_{n,2}} + 1\right)} \tag{21}$$

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

The signal $s_0(t)$ is chosen to be the output of the following system

$$F(s) = \frac{2^6}{(s+2)^6}$$

when an RBS (Random Binary Signal) is injected. We also choose $r(t) = s_0(t)$ for simplicity.

Fig. 2 shows an example of the measured output data $\tilde{y}(t)$ contaminated with noise (noise to signal ratio (NSR) is 30%) when we inject $\boldsymbol{u} = [1, 0, 0, \cdots, 0]^T$ with zero-order hold. We take $\hat{g}_i = \tilde{y}(T_s i)$ $(i = 0, 1, \ldots, N)$ to calculate the estimate of M and $\boldsymbol{\gamma}^0$.

We next proceed to iterative identification. During the iteration process, the measurement noise is white with zero mean and variance σ^2. The variance σ^2 of the measurement noise is chosen so that the NSR will be 30%.

The behavior of the measured output $\tilde{y}(t)$ at the 5th trial is also shown by the dotted line in Fig. 3. The signal $B^k(p)r(t)$ is also shown by the thick line. This figure shows that $y^k(t)$ tracks $B^k(p)r(t)$ very well against the heavy measurement noise, which indicates that the plant model is identified accurately. To confirm this, Bode plots of the estimated system at the 5th trial are shown in Fig. 4 for 50 runs of the algorithm. The thick line represents the Bode plot of the true system.

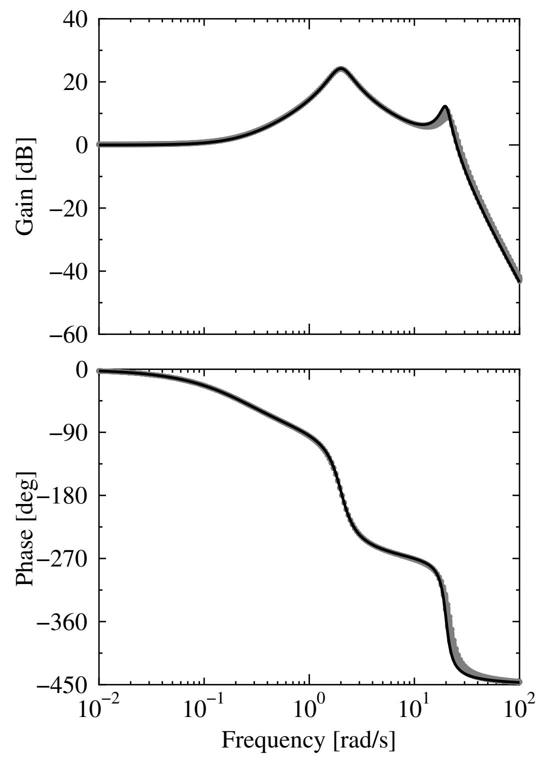

Fig. 4. Bode plots of the estimated system ($k = 5$)

Fig. 5 shows an example of the estimated coefficients at each trial k $(= 1, 2, \cdots, 10)$. From this figure, we see that the initial estimation is improved quickly and the estimation result is robust against measurement noise at each trial. The Euclidean norm of the parameter estimation error is shown in Fig. 6. From this figure, we see that the estimation

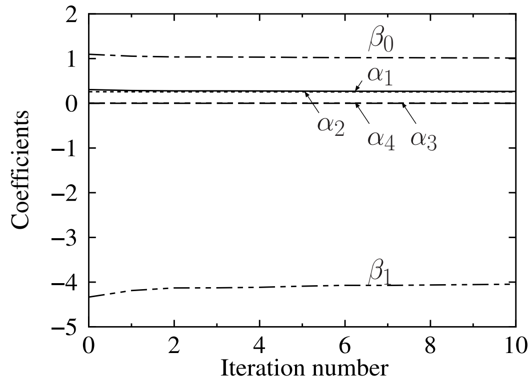

Fig. 5. Identified coefficients γ^k in each trial (True value: $\alpha_1 = 0.26$, $\alpha_2 = 0.255$, $\alpha_3 = 0.003125$, $\alpha_4 = 0.000625$, $\beta_0 = 1$, $\beta_1 = -4$)

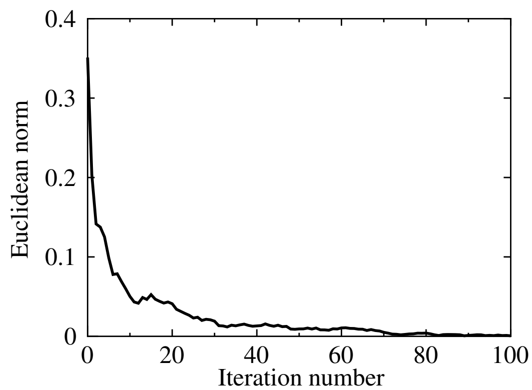

Fig. 6. Euclidean norm of the parameter estimation error $\|\gamma^k - \gamma^\circ\|$

error actually converges to 0 by increasing the number of iterations.

VII. Conclusions

The paper has given a novel approach to identification of linear continuous-time systems directly from the sampled I/O data based on the iterative learning control concept. The method achieves identification through repetition of trials in presence of heavy measurement noise. The robustness against measurement noise is achieved through (i) projection of continuous-time I/O signals onto a finite dimensional parameter space and (ii) Kalman filter type noise reduction. In addition, an alternative simpler method has been given with some robustness analysis. This method has several advantages such as no time-derivatives of I/O data is required and no data re-processing (*e.g.,* decimation or filtering) is necessary. In addition, the identified model quality can be estimated by watching the tracking control performance in experiments. The effectiveness of the proposed method has been demonstrated through numerical examples for a non-minimum phase plant.

References

[1] S. Arimoto. *Control theory of non-linear mechanical systems : A passivity-based and circuit-theoretic approach*, chapter 4 & 5. Oxford Univ. Press, Oxford, 1996.

[2] S. Arimoto, S. Kawamura, and F. Miyazaki. Bettering operation of robotics. *Journal of Robotic System*, 1(2):123–140, 1984.

[3] Z. Bein and J. X. Xu. *Iterative learning control – Analysis, design, integration and applications*. Kluwer Academic Press, Boston, 1998.

[4] Y. Chen and C. Wen. *Iterative learning control: convergence, robustness and applications*, volume LNCIS-248. Springer-Verlag, 1999.

[5] H. Garnier, M. Gilson, and E. Husestein. Developments for the Matlab CONTSID toolbox. In *Proc. of the 13th IFAC Symposium on System Identification, CD-ROM*, 2003.

[6] H. Garnier and M. Mensler. CONTSID: a CONtinuous-Time System IDentification toolbox for matlab. In *Proc. of the 5th European Control Conference*, 1999.

[7] H. Garnier and M. Mensler. The CONTSID toolbox: a matlab toolbox for CONtinuous-Time System IDentification. In *Proc. of the 12th IFAC Symposium on System Identification, CD-ROM*, 2000.

[8] K. Hamamoto and T. Sugie. An iterative learning control algorithm within prescribed input-output subspace. *Automatica*, 37(11):1803–1809, 2001.

[9] K. Hamamoto and T. Sugie. Iterative learning control for robot manipulators using the finite dimensional input subspace. *IEEE Trans. Robotics and Automation*, 18(4):632–635, 2002.

[10] K. L. Moore. *Iterative learning control for deterministic systems*, volume Springer-Verlag Series on Advances in Industrial Control. Springer-Verlag, London, 1993.

[11] G. P. Rao and H. Garnier. Identification of continuous-time systems: Direct or indirect? *Systems Science*, 30(3):25–50, 2004.

[12] F. Miyazaki S. Kawamura and S. Arimoto. Realization of robot motion based on a learning method. *IEEE Trans. on Systems, Man and Cybernetics*, 18(1):126–134, 1988.

[13] F. Sakai and T. Sugie. Continuous-time systems identification based on iterative learning control. In *Proc. of the 16th IFAC World Congress*, 2005.

[14] N. K. Sinha and G. P. Rao. *Identification of Continuous-Time Systems*. Kluwer Academic Publishers, Dordrecht, 1991.

[15] T. Sugie and T. Ono. An iterative learning control law for dynamical systems. *Automatica*, 27(4):729–732, 1991.

[16] T. Sugie and F. Sakai. Noise tolerant iterative learning control for continuous-time systems identification. In *Proc. of the 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005*, pages 4251–4256, 2005.

[17] H. Unbehauen and G. P. Rao. Continuous-time approaches to system identification - a survey. *Automatica*, 26(1):23–35, 1990.

[18] J. X. Xu. The frontiers of iterative learning cotrol ii. *Systems, Control and Information*, 46(2):233–243, 2002.

[19] P. Young. Parameter estimation for continuous-time models - a survey. *Automatica*, 17(1):23–39, 1981.