2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Parameter Estimation Algorithms for Missing-Data Systems

Feng Ding, Jie Ding

Abstract— **This paper considers the problems of parameter identification and output estimation with possibly *irregularly missing* output data, using output error models. By means of an auxiliary (reference) model approach, we present a recursive least squares algorithm to estimate the parameters of missing data systems, and establish convergence properties for the parameter and missing output estimation in the stochastic framework. The basic idea is to replace the unmeasurable inner variables with the output of an auxiliary model.**

I. Introduction

MANY least squares (LS) based identification methods can be used to estimate system parameters, see, e.g., [19], [20], [27]; but most of them assume that input-output measurement data are available at every sampling instant. There are many practical reasons for missing sampled data to arise in system identification, e.g., infrequent/scarce output measurement due to sensory limitation [17], irregularly sampled systems [26], and unexpected interruption in regularly sampled data. In all these cases, standard identification algorithms cannot be applied directly. In this paper, we focus on the problems of parameter identification and output prediction/estimation with *missing output data*; such problems are important because the results can be used not only to monitor the missing output variables, but also to design inferential control schemes with infrequent/scarce output measurements – see the work in [16], [17] for an application in petroleum production.

Another potentially important application of identification with missing data is in network based control systems, in which sensory information from plants is communicated to controllers through network media, and so is the control information to the actuators. Because of the nature of communication networks, packet dropout or data loss is inevitable, yielding systems with missing data.

Due to its practical significance, system identification with missing data has received much attention since 1990s. Isaksson studied identification of ARX models with missing data based on the Kalman filtering (fixed-interval smoothing) technique and off-line maximum likelihood methods [12], whose recursive version was given in [13], but no convergence analysis was carried out. The EM algorithm is a standard tool to develop estimation algorithms in the presence of missing data [9]. Adam *et al.* discussed a parameter estimation problem for ARX models with missing data based on a pseudo-linear regression method [1], but the stability of the predictor used to estimate the missing outputs was not taken into account. Also, Mirsaidi et al studied AR modeling with missing observations [21]. Albertos and co-workers studied output prediction/estimation of a process from scarcely sampled measurements, assuming that the process model was known [3]; however, the convergence analysis of the predictor is based on the assumption of regular output availability – one output measurement for every q input values, namely, a dual-rate sampling pattern. Also, Wallin and Isaksson analyzed stability of the output predictor in [3] based on a state-space formulation, again for regular output availability or a dual-rate sampled case [29]. Finally, Sanchis and Albertos studied the convergence of pseudo-linear recursive algorithms, making the assumption of regular scarce data availability [25].

Missing-data systems with a *regular* pattern can be viewed as dual-rate or multirate sampled-data systems, for which there exists extensive work on control and identification, see, e.g., [2], [4]–[7], [15]–[17], [22], [23], [26]. As dual-rate or multirate systems, such missing-data systems can be treated by the popular lifting technique and converted to time-invariant lifted models to which state-space based identification methods can be extended and from which fast-rate models can be extracted [15]; in addition, a frequency-domain based polynomial transformation technique can be employed to obtain models suitable for identification with slow output sampling (regularly missing output samples) [6], [7]; performance issues related to such estimation problems have been studied in [6] in the stochastic case. However, both the lifting technique and the polynomial transformation technique, as mentioned above, rely on periodicity inherent in regular sampling patterns, and are not suitable for general missing-data systems in which a periodic pattern does not exist. Recently, the authors presented an auxiliary model based least squares algorithm to *directly* identify the parameters of the underlying single-rate models of dual-rate systems, namely regular missing output data cases [8], but the algorithm there cannot be *directly* applied to irregularly missing data cases. The objective of this paper is to extend the auxiliary model approach to study system identification and output estimation problems with *irregular* missing data patterns, and further to analyze convergence properties of the algorithm proposed.

The rest of this paper is organized as follows. In Section II, we introduce the models and describe the estimation problems with missing data; no particular patterns are assumed. In Section III, we propose a parameter and output estimation

This work was supported by the National Natural Science Foundation of China and the Natural Science Foundation of Jiangsu Province (China, BK2007017) and by Program for Innovative Research Team of Jiangnan University.

The authors are with the School of Communication and Control Engineering, Jiangnan University, Wuxi, P. R. China 214122. E-mail addresses: fding@jiangnan.edu.cn; djvsqxb@163.com.

978-1-4244-4524-0/09/$25.00 ©2009 AACC

algorithm. In Section IV, we first introduce some preliminaries, and then present the main results – convergence of both the parameter and output estimation by the proposed algorithm in the stochastic framework. Finally, we offer some concluding remarks in Section VI.

II. MODELS AND DATA MISSING PATTERNS

Let us begin by considering a discrete-time deterministic system with the following input-output relation:

$$y(t) = P_1(z)u(t), \quad P_1(z) = \frac{B(z)}{A(z)}. \tag{1}$$

Here $u(t)$ and $y(t)$ are the system input and output, $P_1(z)$ is the transfer function, z represents the forward shift operator $[z^{-1}u(t) = u(t-1)]$, and $A(z)$ and $B(z)$ are polynomials in z^{-1} defined as:

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a}, \\ B(z) &= b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + \cdots + b_{n_b} z^{-n_b}. \end{aligned}$$

Assume that the orders n_a and n_b are known and define the parameter vector $\boldsymbol{\theta}$ and information vector $\boldsymbol{\varphi}(t)$ as

$$\begin{aligned} \boldsymbol{\theta} &= [a_1, a_2, \cdots, a_{n_a}, b_1, b_2, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\varphi}(t) &= [-y(t-1), -y(t-2), \cdots, -y(t-n_a), \\ &\quad u(t-1), u(t-2), \cdots, u(t-n_b)]^{\mathrm{T}}, \end{aligned}$$

where $n := n_a + n_b$, the superscript T denotes the matrix transpose. Equation (1) may be written in a vector form:

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}. \tag{2}$$

Conventional system identification assumes that the input-output data,

$$\mathcal{U} = \{u(0), u(1), u(2), \cdots\}, \quad \mathcal{Y} = \{y(0), y(1), y(2), \cdots\},$$

are fully available at every sampling instant. In a missing-data system, not the full sets of $\mathcal{U}$ and $\mathcal{Y}$ are available, due to various practical reasons. We can classify missing-data systems into three cases:

1) The *output* missing case: $\mathcal{U}$ is fully available, but only a subset of $\mathcal{Y}$ is available.
2) The *input* missing case: $\mathcal{Y}$ is fully available, but only a subset of $\mathcal{U}$ is available.
3) The *input and output* missing case: Only subsets of $\mathcal{U}$ and $\mathcal{Y}$ are available.

The most common and interesting situation is case 1, because inputs are usually generated by digital computers and are normally available; this case also includes the practically important systems with infrequent and scarce output sampling [17]. If case 2 does arise, it can be converted into a problem in case 1 by inverting the system under some standard assumptions such as stability and minimum phase. Case 3 is considerably different, and is left for the future. In this paper, we will therefore concentrate on case 1, the output missing case.

As in [1], [3], [25], [29], after defining

$$\mu(t) = \begin{cases} 1, & \text{if } y(t) \text{ and } \boldsymbol{\varphi}(t) \text{ are both available,} \\ 0, & \text{otherwise,} \end{cases}$$

the following modified least squares (MLS) algorithm was used to obtain the estimate $\hat{\boldsymbol{\theta}}(t)$ of the unknown parameter vector $\boldsymbol{\theta}$ in (2):

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \mu(t)P(t)\boldsymbol{\varphi}(t)[y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \\ P(t) &= P(t-1) - \mu(t)\frac{P(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)P(t-1)}{1 + \boldsymbol{\varphi}^{\mathrm{T}}(t)P(t-1)\boldsymbol{\varphi}(t)}, \end{aligned}$$

where $P(t)$ is the covariance matrix. This MLS algorithm seems to work in that whenever $y(t)$ or $\boldsymbol{\varphi}(t)$ is unavailable, $\mu(t) = 0$, and hence $\hat{\boldsymbol{\theta}}(t)$ and $P(t)$ are not updated. But it has a major drawback: Even if $y(t)$ is available, $\boldsymbol{\varphi}(t)$ may not be, since $\boldsymbol{\varphi}(t)$ is very likely to contain past missing outputs; this would lead to $\mu(t) = 0$ when $y(t)$ is available. Hence this MLS algorithm does not fully use all available outputs.

Here we argue that the MLS algorithm is not suitable for missing-data systems. Consider a special missing-data system – a dual-rate sampled-data system [2], [6], [7], [16], [17]:

$$y(t) + a_1 y(t-1) + a_2 y(t-2) = b_1 u(t-1) + b_2 u(t-2).$$

Here, all inputs $u(t)$ are available, only outputs $y(0)$, $y(q)$, $y(2q)$, $y(3q)$, $\cdots$, are available ($q \geq 2$ being an integer). We cannot satisfy $\mu(t) = 1$ for any t, because when $t = kq$ (k being an integer), $y(t)$ is available, but the information vector

$$\boldsymbol{\varphi}(t) = [-y(t-1), -y(t-2), u(t-1), u(t-2)]^{\mathrm{T}}$$

always contains missing data. In other words, we cannot formulate any known $\boldsymbol{\varphi}(t)$ using available $y(t)$.

For the output missing case, we define an integer sequence $\{t_s: \ s = 0, 1, 2, \cdots\}$ satisfying

$$0 = t_0 < t_1 < t_2 < t_3 < \cdots < t_{s-1} < t_s < \cdots,$$

with $t_s^* := t_s - t_{s-1} \geq 1$, and assume that $y(t)$ is available only when $t = t_s$ ($s = 0, 1, 2, \cdots$), or equivalently, the set $\{y(t_s): \ s = 0, 1, 2, \cdots\}$ contains all available outputs. For instance, if the available output data are

$$y(0), y(1), y(2), y(5), y(8), y(9), y(10), y(15), \cdots,$$

then we have

$$\begin{aligned} &t_0 = 0, t_1 = 1, t_2 = 2, t_3 = 5, t_4 = 8, t_5 = 9, \\ &t_6 = 10, t_7 = 15, \cdots. \end{aligned}$$

This is a general framework in which we assume no patterns in the output availability; of course, it includes patterned output availability as special cases, e.g., if t_s^* is constant, say, $t_s^* = q$ (a positive integer), we obtain a dual-rate system with input-output sampling ratio equal to q.

III. THE ALGORITHM DESCRIPTION

Consider the stochastic system as shown in Figure 1, with output missing data, described by an output-error model:

$$x(t) = \frac{B(z)}{A(z)}u(t), \quad y(t) = x(t) + v(t), \tag{3}$$

where the inner variable $x(t)$ is the true output (noise-

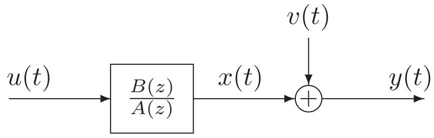

Fig. 1. The output-error system

free output) of the system but unmeasurable; $y(t)$ is the measurable output but is corrupted by the additive noise $v(t)$; the definitions of $A(z)$ and $B(z)$ are as before.

Replacing t in (3) by t_s, it is not difficult to get

$$x(t_s) = \boldsymbol{\varphi}_0^{\mathrm{T}}(t_s)\boldsymbol{\theta}, \quad y(t_s) = \boldsymbol{\varphi}_0^{\mathrm{T}}(t_s)\boldsymbol{\theta} + v(t_s) \tag{4}$$

with

$$\begin{aligned}\boldsymbol{\varphi}_0(t_s) = [&-x(t_s-1), -x(t_s-2), \cdots, -x(t_s-n_a),\\ &u(t_s-1), u(t_s-2), \cdots, u(t_s-n_b)]^{\mathrm{T}}.\end{aligned}$$

Because $x(t_s-i)$ in $\boldsymbol{\varphi}_0(t_s)$ are unknown, the standard least squares algorithm cannot be applied to (4) directly. However, if we replace these unknown $x(t_s-i)$ by their estimates $\hat{x}(t_s-i)$, and $\boldsymbol{\varphi}_0(t_s)$ by $\boldsymbol{\varphi}(t_s)$, then the identification problem of $\boldsymbol{\theta}$ can be solved by using $u(t_s-i)$, $y(t_s)$ and $\hat{x}(t_s-i)$ instead of $x(t_s-i)$. According to the least squares principle, from (4), we propose the following recursive algorithm:

$$\hat{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_{s-1}) + P(t_s)\boldsymbol{\varphi}(t_s)e(t_s), \tag{5}$$

$$e(t_s) = y(t_s) - \boldsymbol{\varphi}^{\mathrm{T}}(t_s)\hat{\boldsymbol{\theta}}(t_{s-1}), \tag{6}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t_{s-1}),\ t \in \{t_{s-1},\ t_{s-1}+1,\ \cdots,\ t_s-1\},$$

$$P^{-1}(t_s) = P^{-1}(t_{s-1}) + \boldsymbol{\varphi}(t_s)\boldsymbol{\varphi}^{\mathrm{T}}(t_s), \quad P(0) = p_0 I, \tag{7}$$

$$\hat{x}(t_s-j) = \boldsymbol{\varphi}^{\mathrm{T}}(t_s-j)\hat{\boldsymbol{\theta}}(t_{s-1}), \quad j = 1,\ 2,\ \cdots,\ t_s^*, \tag{8}$$

$$\begin{aligned}\boldsymbol{\varphi}(t_s-j) = [&-\hat{x}(t_s-1-j), \cdots, -\hat{x}(t_s-n_a-j),\\ &u(t_s-1-j), \cdots, u(t_s-n_b-j)]^{\mathrm{T}},\end{aligned} \tag{9}$$

where $\hat{\boldsymbol{\theta}}(t)$ represents the estimate of $\boldsymbol{\theta}$, and p_0 is a large positive real number. As in many references [1], [3], [25], [29], we simply hold $\hat{\boldsymbol{\theta}}(t)$ unchanged on the interval $[t_{s-1}+1, t_s-1]$ for which outputs are unavailable.

This identification algorithm is based on output error models instead of ARX models as in [1], [12] for the missing data systems, and differs from those mentioned in the introduction in that it directly estimate the true output (noise-free output) $x(t)$ instead of unavailable $y(t)$ as in [1], [3], [6], [25], [29] using the available data.

IV. Main results

In this section, we present and prove the main convergence results on the parameter and output estimation with the algorithm proposed earlier.

Let us begin by defining some notation. The symbols $\lambda_{\max}[X]$ and $\lambda_{\min}[X]$ represent the maximum and minimum eigenvalues of the symmetric matix X, respectively; $|X|$ denotes the matrix determinant, and $\|X\|^2 = \mathrm{tr}[XX^{\mathrm{T}}]$, the trace of XX^{T}. For $g(t) \geq 0$, we write $f(t) = O(g(t))$ if there exists finite positive constants δ_1 and t_0 such that $|f(t)| \leq \delta_1 g(t)$ for $t \geq t_0$, and $f(t) = o(g(t))$ if $f(t)/g(t) \to 0$ as $t \to \infty$.

Define

$$P_0^{-1}(t_s) := P_0^{-1}(t_{s-1}) + \boldsymbol{\varphi}_0(t_s)\boldsymbol{\varphi}_0^{\mathrm{T}}(t_s), \quad P_0(0) = p_0 I;$$

$$r(t_s) := \mathrm{tr}[P^{-1}(t_s)]; \quad r_0(t_s) := \mathrm{tr}[P_0^{-1}(t_s)].$$

Hence, we easily get

$$|P^{-1}(t_s)| \leq r^n(t_s); \quad r(t_s) \leq n\lambda_{\max}[P^{-1}(t_s)]; \tag{10}$$

Define the parameter estimation error vector $\tilde{\boldsymbol{\theta}}(t_s)$ and a nonnegative definite function $V(t_s)$ as

$$\tilde{\boldsymbol{\theta}}(t_s) = \hat{\boldsymbol{\theta}}(t_s) - \boldsymbol{\theta}, \tag{11}$$

$$V(t_s) = \tilde{\boldsymbol{\theta}}^{\mathrm{T}}(t_s)P^{-1}(t_s)\tilde{\boldsymbol{\theta}}(t_s). \tag{12}$$

Theorem 1: For the missing-data system in (4) and the algorithm in (5)-(9), assume that $\{v(t)\}$ is statistically independent of the input $u(t)$ and satisfies

$$(A1) \quad \mathrm{E}[v(t)|\mathcal{F}_{t-1}] = 0,\ \text{a.s.},$$

$$(A2) \quad \mathrm{E}[v^2(t)|\mathcal{F}_{t-1}] \leq \sigma_v^2[\ln|P^{-1}(t)|]^{\varepsilon_1},\ \text{a.s.},$$

where $\sigma_v^2 < \infty$, $0 \leq \varepsilon_1 < \infty$, $\{\mathcal{F}_t\}$ is the σ-algebra sequence generated by the observations up to and including time t [10]. Suppose that $A(z)$ is stable (all roots of $A(z)$ are inside the unit circle), and that there exist positive constants c, α, β and k such that for $s \geq k$, the following generalized persistent excitation condition (unbounded condition number) holds:

$$(C1) \quad \alpha I \leq \frac{1}{s}\sum_{i=1}^{s}\boldsymbol{\varphi}_0(t_i)\boldsymbol{\varphi}_0^{\mathrm{T}}(t_i) \leq \beta s^c I,\ \text{a.s.}$$

Then for any ε with $0 \leq \varepsilon_1 < \varepsilon < \infty$, the parameter estimation error associated with the algorithm in (5)-(9) satisfies:

$$\|\hat{\boldsymbol{\theta}}(t_s) - \boldsymbol{\theta}\|^2 = O\left(\frac{[\ln s]^{1+\varepsilon}}{s}\right) \to 0,\ \text{a.s.},\ as\ s \to \infty.$$

Theorem 1 shows that the parameter estimation error $\|\hat{\boldsymbol{\theta}}(t_s) - \boldsymbol{\theta}\|^2$ converges to zero at the rate of $O\left(\frac{[\ln s]^{1+\varepsilon}}{s}\right)$ ($0 \leq \varepsilon_1 < \varepsilon < \infty$): A high noise level results in a slow rate of convergence; in other words, the convergence rate becomes slower for large ε_1.

Theorem 2: For the missing-data system in (4) and the algorithm in (5)-(9), assume that $\{v(t)\}$ is a random noise sequence with zero mean, and is statistically independent of $u(t)$ [hence (A1) holds], and furthermore

$$\begin{aligned}(A3) \quad &\mathrm{E}[v^2(t)|\mathcal{F}_{t-1}] \leq \sigma_v^2 r^{\epsilon_1}(t),\ \text{a.s.},\ 0 \leq \sigma_v^2 < \infty,\\ &0 \leq \epsilon_1 < 1.\end{aligned}$$

Suppose $A(z)$ is stable, and that the weak persistent excitation (WPE) condition (bounded condition number) holds, i.e., take $c = 0$ in (C1) to get

$$\begin{aligned}(C2) \quad &\alpha I \leq \frac{1}{s}\sum_{i=1}^{s}\boldsymbol{\varphi}_0(t_i)\boldsymbol{\varphi}_0^{\mathrm{T}}(t_i) \leq \beta I,\ \text{a.s.},\\ &0 < \alpha \leq \beta < \infty\ and\ large\ s.\end{aligned}$$

Then for any ϵ with $0 \leq \epsilon_1 < \epsilon < 1$, the estimation error satisfies

$$\|\hat{\boldsymbol{\theta}}(t_s) - \boldsymbol{\theta}\|^2 = O\left(\frac{1}{s^{1-\epsilon}}\right) \to 0, \text{ a.s., } as\ s \to \infty.$$

The proofs of Theorems 1 to 4 are omitted in order to save space but available from the authors.

Theorem 2 indicates that the parameter estimation error $\|\hat{\boldsymbol{\theta}}(t_s) - \boldsymbol{\theta}\|^2$ converges to zero at the rate of $O\left(\frac{1}{s^{1-\epsilon}}\right)$ $(0 \leq \epsilon_1 < \epsilon < 1)$; a smaller ϵ_1 (a lower noise level) will lead to a faster rate for the parameter estimates to converge to the true parameters.

In the area of identification algorithms, Ljung's and Solo's consistency analysis was based on the condition that the input and output signals have finite nonzero power, and assumed that the noise is independent and identically distributed random sequence with finite 4th-order moments [18], or the process noise and input are stationary and ergodic [28]. Also, Lai and Wei [14], Guo and Chen [11], and Ren and Kumar [24] assumed that the high-order moments of the noise $\{v(t)\}$ exist, i.e., $\mathrm{E}[v^\gamma(t)|\mathcal{F}_{t-1}] < \infty,$ a.s. for some $\gamma > 2$. Notice that such assumptions have not been made in our results.

Most identification algorithms of missing data systems assume that the models considered are ARX models, but few consider the output error models with missing data. For the output error models with missing data, a difficulty arises in that the information vector contains unknown noise-free (true) outputs $x(\cdot)$. To this point, we replace these unknown $x(\cdot)$ by using the estimates $\hat{x}(\cdot)$ computed by an auxiliary model in (8).

Theorem 3: For the system in (4), assume the conditions stated in Theorem 1 are satisfied; furthermore, let $\Delta_s := \max[t_1^*, t_2^*, \cdots, t_s^*]$ and assume that there exist constants $0 < \delta_1 < \infty$ and $0 \leq \mu < 1$ such that the following conditions hold:

$$(A4) \quad t_{s+1} \leq \delta_1 t_s,\ or\ t_{s+1} = O(t_s),$$
$$(A5) \quad \Delta_s \leq \delta_1 t_s^\mu,\ or\ \Delta_s = O(t_s^\mu).$$

Then the bounded input assumption implies that there exists a positive integer k such that the output estimation error, $\hat{x}(t) - x(t)$, satisfies:

$$1. \quad \sum_{i=t_k}^{t_s} [\hat{x}(i) - x(i)]^2 = O(t_s^\mu [\ln t_s]^{2+\varepsilon}), \text{ a.s.;}$$

$$2. \quad \frac{1}{t}\sum_{i=1}^{t} [\hat{x}(i) - x(i)]^2 = O\left(\frac{[\ln t]^{2+\varepsilon}}{t^{1-\mu}}\right) \to 0, \text{ a.s.}$$

Note the following:

- Conditions (A4) and (A5) are reasonable and common assumptions and include the cases in which the missing data length between two successive available measurements is finite, i.e., $\mu = 0$.
- Theorem 3 indicates that the output estimation error $[\hat{x}(t) - x(t)]^2$ converges to zero in the average sense; a smaller μ and ε (i.e., a lower noise level) will lead to a faster convergence rate.

Furthermore, the following result on output estimation can be established.

Theorem 4: For the system in (4), if the conditions of Theorem 2 and (A4)-(A5) hold, and $\mu + \epsilon < 1$, then the bounded input assumption implies that there exists a positive integer k such that the output estimation error satisfies:

$$1. \quad \sum_{i=t_k}^{t_s} [\hat{x}(i) - x(i)]^2 = O(t_s^{\mu+\epsilon}), \text{ a.s.;}$$

$$2. \quad \frac{1}{t}\sum_{i=1}^{t} [\hat{x}(i) - x(i)]^2 = O\left(\frac{1}{t^{1-\mu-\epsilon}}\right) \to 0, \text{ a.s.}$$

From Theorems 3 and 4, we can arrive at the following conclusions: Once the parameter estimates converge to the true parameters, $\hat{x}(t)$ is the best estimator for $x(t)$. If $v(t)$ is assumed to be white, then $\hat{x}(t)$ can be viewed as the optimal estimate for $y(t)$.

V. Example

Example Consider the system:

$$y(t) = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} u(t) + v(t)$$
$$= \frac{0.40 z^{-1} + 0.30 z^{-2}}{1 - 1.60 z^{-1} + 0.80 z^{-2}} u(t) + v(t).$$

$\{u(t)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance $\sigma_u^2 = 1.00^2$, and $\{v(t)\}$ as a white noise sequence with zero mean and variance σ_v^2. We assume that $\{u(t), y(t_s)\}$ are measured for $t_s^* = 1$ (no missing data) and $t_s^* = 3$ (2 missing samples for every 3 output samples), and apply the proposed algorithm to estimate the parameters (a_i, b_i) of this system. The parameter estimates and their errors are shown in Tables I and II, and the parameter estimation errors versus s are shown in Figure 2, where δ_{ns} is the noise-to-signal ratio defined by the square root of the ratio of the variances of $v(t)$ and $x(t)$, i.e.,

$$\delta_{\text{ns}} = \sqrt{\frac{\text{var}[v(t)]}{\text{var}[x(t)]}} \times 100\% = \frac{\sigma_v}{\sigma_x} \times 100\%,$$

$\delta = \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ is the relative parameter estimation error, $\sigma_v^2 = 1.00^2$, and $\delta_{\text{ns}} = 40.38\%$.

From Tables I and II, as well as Figure 2, it is clear that the δ's are becoming smaller (in general) as s increases. This validates the proved results in the earlier sections.

VI. Conclusions

A recursive least squares algorithm is presented for missing-data systems with no regular patterns; the algorithm estimates simultaneously the system parameters and unknown outputs. Convergence performance of the proposed estimation algorithm is analyzed in detail in the stochastic framework. The algorithm is also evaluated by simulated examples. The methods developed can be extended to identify missing-data systems with colored noises. Since the estimate $\hat{x}(t)$ consistently converges to the true (noise-free) output

TABLE I

THE PARAMETER ESTIMATES ($t_s^* = 1$, $\sigma_v^2 = 1.00^2$, $\delta_{\rm ns} = 40.38\%$)

s	a_1	a_2	b_1	b_2	δ (%)
100	-1.19129	0.43371	0.61319	0.82247	42.37987
200	-1.20280	0.41630	0.48995	0.61344	34.52910
300	-1.26393	0.45255	0.49033	0.54822	29.65696
500	-1.41130	0.59023	0.44051	0.35467	15.62599
1000	-1.64057	0.80729	0.37221	0.20757	5.65048
1500	-1.62765	0.82301	0.37947	0.23680	4.06808
2000	-1.63625	0.83488	0.38370	0.25339	3.79510
2500	-1.62835	0.83332	0.40147	0.25448	3.39980
3000	-1.60902	0.81172	0.40500	0.27014	1.81378
3500	-1.60898	0.80891	0.40675	0.27637	1.48840
4000	-1.60501	0.80607	0.40665	0.28114	1.15739
True values	-1.60000	0.80000	0.40000	0.30000	

TABLE II

THE PARAMETER ESTIMATES ($t_s^* = 3$, $\sigma_v^2 = 1.00^2$, $\delta_{\rm ns} = 40.38\%$)

s	a_1	a_2	b_1	b_2	δ (%)
100	-0.33104	-0.23949	-0.04825	0.66703	93.66076
200	-0.73688	-0.16743	0.14208	0.73214	74.87500
300	-0.77237	-0.07714	0.24124	0.61706	67.67520
500	-0.96692	0.17496	0.44221	0.57158	50.13012
1000	-1.20677	0.38942	0.38922	0.54031	33.23465
1500	-1.45052	0.62712	0.40057	0.41019	13.66000
2000	-1.54027	0.72650	0.38557	0.37842	6.66546
2500	-1.58090	0.78081	0.37351	0.35319	3.51563
3000	-1.58869	0.79202	0.36809	0.35545	3.52395
3500	-1.58490	0.78656	0.38151	0.35116	3.12444
4000	-1.58491	0.78960	0.38737	0.34701	2.80019
True values	-1.60000	0.80000	0.40000	0.30000	

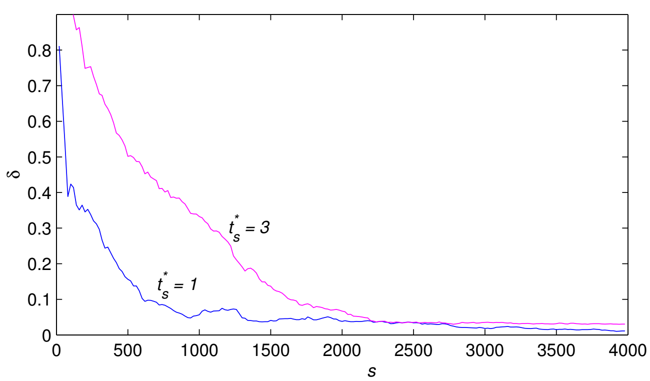

Fig. 2. The parameter estimation errors vs. s ($t_s^* = 1, 3$)

$x(t)$, it is also potentially useful as a basis for inferential control of missing-data systems by using $\hat{x}(t)$ in feedback loops [16].

REFERENCES

[1] G.J. Adams, P. Albertos, G.C. Goodwin and A.J. Isaksson, "Parameter estimation for ARX models with missing data," *Proc. IFAC Symposium on System Identification (SYSID'94)*, July 4-6, 1994, Copenhagen, Denmark, pp. 163-168.

[2] P. Albertos, J. Salt and J. Tormero, "Dual-rate adaptive control," *Automatica*, vol. 32. no. 7, pp. 1027-1030, 1996.

[3] P. Albertos, R. Sanchis and A. Sala, "Output prediction under scarce data operation: control applications," *Automatica*, vol. 35, no. 10, pp. 1671-1681, 1999.

[4] T. Chen and B.A. Francis, *Optimal Sampled-data Control Systems*. London: Springer, 1995.

[5] T. Chen and L. Qiu, "$\mathcal{H}_\infty$ Design of general multirate sampled-data control systems," *Automatica*, vol. 30, no. 7, pp. 1139-1152, 1994.

[6] F. Ding, X.P. Liu, Y.Shi, "Convergence analysis of estimation algorithms of dual-rate stochastic systems," *Applied Mathematics and Computation*, vol. 176, no. 1, pp. 245-261, 2006.

[7] F. Ding and T. Chen, "Parameter estimation for dual-rate systems with finite measurement data," *Dynamics of Continuous, Discrete and Impulsive Systems, Series B: Applications & Algorithms*, vol. 11, no. 1, pp. 101-121, 2004.

[8] F. Ding and T. Chen, "Combined parameter and output estimation of dual-rate systems using an auxiliary model," *Automatica*, vol. 40, no. 10, pp. 1739-1748, 2004.

[9] G. C. Goodwin and A. Feuer, "Estimation with missing data," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 5, no.3, pp. 220C244, 1998.

[10] G.C. Goodwin and K.S. Sin, *Adaptive Filtering, Prediction and Control*, Englewood Cliffs, New Jersey: Prentice-Hall, 1984.

[11] L. Guo and H.F. Chen, "The Åström-Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers," *IEEE Trans. Automat. Contr.*, vol. 36, no. 7, pp. 802-812, 1991.

[12] A.J. Isaksson, "Identification of ARX-models subject to missing data," *IEEE Trans. Automat. Contr.*, vol. 38, no. 5, pp. 813-819, 1993.

[13] A.J. Isaksson, "A recursive EM algorithm for identification subject to missing data," *Proc. IFAC Symposium on System Identification (SYSID'94)*, July 4-6, 1994, Copenhagen, Denmark, pp. 953-958.

[14] T.L. Lai and C.Z. Wei, "Extended least squares and their applications to adaptive control and prediction in linear systems," *IEEE Trans. Automat. Contr.*, vol. 31, no. 10, pp. 898-906, 1986.

[15] D. Li, S.L. Shah and T. Chen, "Identification of fast-rate models from multirate data," *International Journal of Control*, vol. 74, no. 7, pp. 680-689, 2001.

[16] D. Li, S.L. Shah, and T. Chen, "Analysis of dual-rate inferential control systems," *Automatica*, vol. 38, no. 6, pp. 1053-1059, 2002.

[17] D. Li, S.L. Shah, T. Chen, and K.Z. Qi, "Application of dual-rate modeling to CCR octane quality inferential control," *IEEE Trans. Contr. Syst. Technol.*, vol. 11, no. 1, pp. 43-51, 2003.

[18] L. Ljung, "Consistency of the least-squares identification method," *IEEE Trans. Automat. Contr.*, vol. 21, no. 5, pp. 779-781, 1976.

[19] L. Ljung and T. Södeström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[20] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall, 1999.

[21] S. Mirsaidi, G.A. Fleury and J. Oksman, "LMS-like AR modeling in the case of missing observations," *IEEE Trans. on Signal Processing*, vol. 45, no. 6, pp. 1574-1583, 1997.

[22] L. Qiu and T. Chen, "$\mathcal{H}_\infty$-optimal design of multirate sampled-data systems," *IEEE Trans. Automat. Contr.*, vol. 39, no. 12, pp. 2506-2511, 1994.

[23] L. Qiu and T. Chen, "Multirate sampled-data systems: All $\mathcal{H}_\infty$ suboptimal controllers and the minimum entropy controllers," *IEEE Trans. Automat. Contr.*, vol. 44, no. 3, pp. 537-550, 1999.

[24] W. Ren and P.K. Kumar, "Stochastic adaptive prediction and model reference control," *IEEE Trans. Automat. Contr.*, vol. 39, no. 10, pp. 2047-2060, 1994.

[25] R. Sanchis and P. Albertos, "Recursive identification under scarce measurements: convergence analysis," *Automatica*, vol. 38, no. 3, pp. 535-544, 2002.

[26] J. Sheng, T. Chen and S.L. Shah, "Generalized predictive control for non-uniformly sampled systems," *Journal of Process Control*, vol. 12, no. 8, pp. 875-88, 2002.

[27] T. Söderström and P. Stoica, *System Identification*. Englewood-Cliffs, New Jersey: Prentice-hall, 1988.

[28] V. Solo, "The convergence of AML," *IEEE Trans. Automat. Contr.*, vol. 24, no. 6, 958-962, 1979.

[29] R. Wallin, A.J. Isaksson and O. Noréus, "Extensions to output prediction under scarce data operation: control applications," *Automatica*, vol. 37, no. 12, pp. 2069-2071, 2001.