

Parameter Estimation with Missing Input/Output Data

Huazhen Fang, Yang Shi, and Jian Wu

Abstract—The problem of recursive parameter estimation with missing input/output data is studied in this paper. A fictitious measurement noise model is presented for missing data, and a noise-robust minimum component analysis based algorithm is developed to recursively estimate parameters from the new ‘noisy’ input/output data. Convergence properties of the proposed algorithm are analyzed. The simulation results verify the effectiveness of the proposed algorithm.

I. INTRODUCTION

System parameter estimation is an active research field in signal processing, spectral analysis and modern control. Many well developed methods have been proposed, such as least squares (LS), stochastic gradients (SG), maximum likelihood (ML), etc. A common premise for these methods is that all the input and output data are available for use. However, the real-life applications may be subject to various forms of missing data. For instance, when an irregularly or infrequently or multirate sampling scheme is used or sensor failure happens, ‘missing’ measurements would appear. Recently, networked control systems, whose control loops are physically constructed on communication networks, has attracted much interest [1], [2], [3]. In such systems, data transmission often suffers from time delay or even loss, which can be potentially modeled as data missing. In fact, parameter estimation with missing data is a fundamental challenge in practice, and has thus far received remarkable attention.

Most current works consider only output missing, and a widely used scheme is based on output estimation plus parameter estimation. That is, the missing outputs are estimated first, and then some classical methods like LS and ML are applied on the reconstructed data set for parameter estimation, see [4], [5], [6], [7], [8]. However, the problem becomes more complicated when both inputs and outputs are affected by missing values. In [9], a frequency domain solution was developed by treating the missing data as unknown parameters. Chen *et al.* proposed an iterative least squares (IRLS) based algorithm that minimizes l_p norm estimation errors [10]. An IRLS technique was also discussed in [11], but it was developed from matrix manipulations. It is noted that, in the case of missing data in both inputs and outputs, 1) no recursive method has ever been proposed, though recursion is crucial for most adaptive schemes in filtering, prediction and control and 2) no effective method

has been developed for parameter estimation subject to a large amount of missing data.

The objective of this paper is to construct a recursive parameter estimation algorithm capable of handling a great deal of data loss. In our approach, the missing input/output values are neatly modeled as the consequences of measurement noises. It is known that minimum component analysis (MCA) is strongly noise-robust when applied to parameter estimation. Therefore, recursive MCA (RMCA) is promisingly used to identify parameters from the available input/output data.

The rest of the paper is organized as follows: In Section II, we introduce the models and state the problem of interest. Section III gives an RMCA based algorithm. A numerical example is studied in Section IV to show the effectiveness of the algorithm proposed. Finally, some concluding remarks are offered in Section V.

II. PROBLEM FORMULATION

Let us begin by considering a deterministic linear discrete-time system:

$$y_o(t) = \frac{B(z)}{A(z)}u_o(t). \quad (1)$$

Here $u_o(t)$ and $y_o(t)$ represent the system input and output, and $A(z)$ and $B(z)$ are two polynomials in z^{-1} (backshift operator):

$$\begin{aligned} A(z) &= 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a}, \\ B(z) &= b_0 + b_1z^{-1} + b_2z^{-2} + b_3z^{-3} + \dots + b_{n_b}z^{-n_b}. \end{aligned}$$

Assume that the orders n_a and n_b are known. When measured, $u_o(t)$ and $y_o(t)$ are corrupted by additive noises $w(t)$ and $v(t)$, respectively:

$$u_m(t) = u_o(t) + w(t), \quad (2)$$

$$y_m(t) = y_o(t) + v(t). \quad (3)$$

Suppose that some of the input and output measurements will be missing randomly. At time instant t , the availability of $u_m(t)$ and $y_m(t)$ relies on the binary random variables $\gamma(t)$ and $\lambda(t)$, respectively, that is,

$$\begin{aligned} u(t) &= \gamma(t)u_m(t), \\ y(t) &= \lambda(t)y_m(t), \end{aligned}$$

with

$$\begin{aligned} \gamma(t) &= \begin{cases} 1, & \text{if } u_m(t) \text{ is available,} \\ 0, & \text{else if } u_m(t) \text{ is unavailable,} \end{cases} \\ \lambda(t) &= \begin{cases} 1, & \text{if } y_m(t) \text{ is available,} \\ 0, & \text{else if } y_m(t) \text{ is unavailable.} \end{cases} \end{aligned}$$

This work was supported by the Natural Sciences and Engineering Research Council of Canada and Canadian Foundation of Innovation.

H. Fang, Y. Shi and Jian Wu are with the Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, S7N 5A9, Canada. {huazhen.fang, yang.shi, jian.wu}@usask.ca

Clearly, we have

$$u(t) = \gamma(t) [u_o(t) + w(t)], \quad (4)$$

$$y(t) = \lambda(t) [y_o(t) + v(t)]. \quad (5)$$

Define the parameter vector and regression vectors as follows:

$$\theta = [a_1 \ a_2 \ \cdots \ a_{n_a} \ b_1 \ b_2 \ \cdots \ b_{n_b}]^T.$$

Note that θ is actually not unique. Any $\rho\theta$ for some non-zero and finite constant ρ would produce identical input and output measurements. In other words, any identification scheme cannot distinguish between θ and $\rho\theta$. Therefore, without loss of generality, we fix the value of θ by imposing the constraint: 1) $\theta^T\theta = 1$; 2) the first element of θ is positive.

We make the following assumptions throughout the paper:

- A1. The input sequence $\{u_o(t)\}$ is persistently excited, and is a stationary ergodic random sequence.
- A2. The sequences $\{\gamma(t)\}$ and $\{\lambda(t)\}$ are independent of $u_m(t)$ and $y_m(t)$, respectively, and mutually independent.
- A3. The probabilities of occurrence of missing data for both inputs and outputs are equal. Let

$$\begin{aligned} P\{\gamma(t), \lambda(t) = 1\} &= p, \\ P\{\gamma(t), \lambda(t) = 0\} &= 1 - p, \end{aligned}$$

where p is known.

- A4. The noise sequences $\{w(t)\}$ and $\{v(t)\}$ are white Gaussian noises, and independent of $u_o(t)$ and $y_o(t)$, respectively. Their variances are assumed known.

In what follows, we shall study how to recursively estimate the parameter vector θ from the available inputs and outputs, i.e., $\{u(t)\}$ and $\{y(t)\}$.

III. MAIN RESULTS

This section models data missing and rebuilds the system first, followed by RMCA based parameter estimation. A summary of the final algorithm is provided in the end.

A. Data Missing Modeling and System Rebuilding

When missing data occurs, a ‘let-it-be’ strategy is adopted: regard missing values as the counteracting effects of fictitious measurement noises and just replace them by 0. Take the input for example. At time instant t , if input missing happens, i.e., $\gamma(t) = 0$, we suppose that the actual $u_m(t)$ is counteracted by a measurement noise with amplitude of $-u_m(t)$. As a result, we get $u(t) = 0$. This strategy is depicted in Fig. 1. It can also be applied identically to output missing.

Mathematically, (4) and (5) can be rewritten as

$$u(t) = u_o(t) + \omega(t), \quad (6)$$

$$y(t) = y_o(t) + v(t), \quad (7)$$

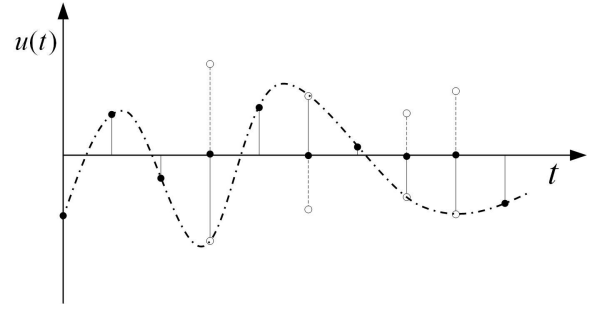


Fig. 1: The fictitious noise model for data missing (The empty circles represent missing data and the dashed-line segments denote the fictitious measurement noises.)

where

$$\begin{aligned} \omega(t) &= [\gamma(t) - 1]u_o(t) + \gamma(t)w(t) \\ &= [\gamma(t) - 1]u_m(t) + w(t), \\ v(t) &= [\lambda(t) - 1]y_o(t) + \lambda(t)v(t) \\ &= [\lambda(t) - 1]y_m(t) + v(t). \end{aligned}$$

Note the similarity between (2)-(3) and (6)-(7). It indicates that $\omega(t)$ and $v(t)$ can be treated as new measurement noises in the considered system, which matches the ‘let-it-be’ strategy.

As will be seen later, it is necessary to unify $\text{Var}\{\omega(t)\}$ and $\text{Var}\{v(t)\}$. It is not difficult to derive that

$$\text{Var}\{\omega(t)\} = (1 - p)\text{Var}\{u_m(t)\} + \text{Var}\{w(t)\}, \quad (8)$$

$$\text{Var}\{v(t)\} = (1 - p)\text{Var}\{y_m(t)\} + \text{Var}\{v(t)\}. \quad (9)$$

Let ε be the variance ratio:

$$\varepsilon = \sqrt{\frac{\text{Var}\{\omega(t)\}}{\text{Var}\{v(t)\}}}. \quad (10)$$

Then it holds that,

$$\text{Var}\{\varepsilon v(t)\} = \text{Var}\{\omega(t)\}.$$

Multiplying both sides of the system (1) by ε leads to an equivalent system:

$$\varepsilon y_o(t) = \frac{\varepsilon B(z)}{A(z)} u_o(t). \quad (11)$$

Based on the new system (11), we define:

$$\begin{aligned} \vartheta &= [a_0 \ a_1 \ \cdots \ a_{n_a} \ \varepsilon b_0 \ \varepsilon b_1 \ \cdots \ \varepsilon b_{n_b}]^T, \\ \psi(t) &= [-\varepsilon y(t) \ -\varepsilon y(t-1) \ \cdots \ -\varepsilon y(t-n_a) \\ &\quad u(t) \ u(t-1) \ \cdots \ u(t-n_b)]^T. \end{aligned}$$

From ϑ , we can easily reach θ if ε is known. Thus it will be our focus to estimate ϑ from the regression vectors $\psi(t)$ s. Before doing that, we need to design a numerical way to calculate ε .

However, note that the theoretical value of ε cannot be obtained directly, and it should be calculated *a priori* by

sample averaging. Using the first r sample points, $\text{Var}\{\omega(t)\}$ and $\text{Var}\{v(t)\}$ can be unbiasedly estimated:

$$\widehat{\text{Var}}\{\omega(t)\} = \frac{1-p}{p} \cdot \frac{1}{r-1} \sum_{t=1}^r \left[u(t) - \frac{1}{r} \sum_{t=1}^r u(t) \right]^2 + \sigma_w^2, \quad (12)$$

$$\widehat{\text{Var}}\{v(t)\} = \frac{1-p}{p} \cdot \frac{1}{r-1} \sum_{t=1}^r \left[y(t) - \frac{1}{r} \sum_{t=1}^r y(t) \right]^2 + \sigma_v^2. \quad (13)$$

Further, ε is estimated by

$$\hat{\varepsilon} = \sqrt{\frac{\widehat{\text{Var}}\{\omega(t)\}}{\widehat{\text{Var}}\{v(t)\}}}. \quad (14)$$

B. The Algorithm

MCA is a very useful approach to parameter estimation. Huang, for the first time, proposes batch-wise MCA based parameter estimation methods [12]. It is also shown in [12] that the method is robust against input/output measurement noises.

Consider the batch-wise case first. Collect all the regression vectors $\psi(t)$ for $r < t \leq N$, and calculate the empirical mean vector of $\psi(t)$ s:

$$\bar{\psi} = \frac{1}{N-r} \sum_{t=r+1}^N \psi(t).$$

The covariance matrix is then given by

$$R = \frac{1}{N-r-1} \sum_{t=r+1}^N [\psi(t) - \bar{\psi}] [\psi(t) - \bar{\psi}]^T. \quad (15)$$

Finally, eigendecompose R to obtain its eigenvalues and eigenvectors. Denote \underline{q} the minimum component (MC) of R , i.e., the eigenvector associated to the minimum eigenvalue. Actually \underline{q} is the very estimate of ϑ , denoted by $\hat{\vartheta}$. Dividing the last n_b elements of $\hat{\vartheta}$ by ε and combining them with the first n_a elements give the estimate of θ , $\hat{\theta}$. The above is the basic procedures of MCA.

As pointed out in Theorem 5 of [12], for the system (11) that has unified input/output measurement variances, $\hat{\vartheta}$ yielded by the batch-wise MCA is a consistent estimate of ϑ .

For the above batch-wise MCA based parameter estimation with missing input/output data, its recursive counterpart is ready to be derived. We would update and eigendecompose R when each new $\psi(t)$ arrives.

Define

$$\begin{aligned} \bar{\psi}(t) &= \frac{1}{t} \sum_{i=1}^t \psi(i) \\ &= \frac{t-1}{t} \bar{\psi}(t-1) + \frac{1}{t} \psi(t), \end{aligned} \quad (16)$$

$$\begin{aligned} \Delta \bar{\psi}(t) &= \bar{\psi}(t) - \bar{\psi}(t-1) \\ &= -\frac{1}{t} \bar{\psi}(t-1) + \frac{1}{t} \psi(t). \end{aligned} \quad (17)$$

Then from (15) it follows that

$$\begin{aligned} R(t) &= \frac{t-1}{t} R(t-1) + \frac{t-1}{t} \Delta \bar{\psi}(t) \Delta \bar{\psi}^T(t) \\ &\quad + \frac{1}{t} (\psi(t) - \bar{\psi}(t)) (\psi(t) - \bar{\psi}(t))^T, \end{aligned} \quad (18)$$

which is two rank-one modifications [13]. Suppose $R(t) = Q(t)\Lambda(t)Q^T(t)$, where $Q(t)$ is the orthonormal eigenvector matrix and $\Lambda(t)$ is the diagonal eigenvalue matrix. With such notation, (18) becomes

$$\begin{aligned} Q(t)t\Lambda(t)Q^T(t) &= Q(t-1)[(t-1)\Lambda(t-1) \\ &\quad + \alpha(t)\alpha^T(t) + \beta(t)\beta^T(t)] Q^T(t-1), \end{aligned}$$

where

$$\alpha(t) = \sqrt{t-1} Q^T(t-1) \Delta \bar{\psi}(t), \quad (19)$$

$$\beta(t) = Q^T(t-1) (\psi(t) - \bar{\psi}(t)). \quad (20)$$

Further by doing eigendecomposition

$$\begin{aligned} [(t-1)\Lambda(t-1) + \alpha(t)\alpha^T(t) + \beta(t)\beta^T(t)] \\ = V(t)D(t)V^T(t), \end{aligned}$$

we have

$$Q(t)t\Lambda(t)Q^T(t) = Q(t-1)V(t)D(t)V^T(t)Q^T(t-1).$$

By direct observation, it is found that

$$Q(t) = Q(t-1)V(t), \quad (21)$$

$$\Lambda(t) = \frac{1}{t} D(t). \quad (22)$$

After each eigenpair, $Q(t)$ and $\Lambda(t)$, is obtained, we can easily find the MC of $R(t)$, $\underline{q}(t)$, which is also $\hat{\vartheta}(t)$, the estimate of ϑ at time instant t . With ε calculated *a priori*, $\hat{\theta}(t)$ can be obtained from $\hat{\vartheta}(t)$.

The final algorithm is outlined as follows:

- 1: Collect $u(t)$ and $y(t)$ for $1 \leq t \leq r$
- 2: Calculate an approximate ε [Equations (10) and (12)-(14)]
- 3: Calculate $\bar{\psi}(r)$ and $\Delta \bar{\psi}(r)$
- 4: Initialize $Q(0)$ and $\Lambda(0)$
- 5: **for** $t = r+1$ to N **do**
- 6: Get the regression vector $\psi(t)$
- 7: Calculate $\bar{\psi}(t)$ and $\Delta \bar{\psi}(t)$ [Equations (16)(17)]
- 8: Update $\alpha(t)$ and $\beta(t)$ [Equations (19)(20)]
- 9: Eigendecompose

$$[(t-1)\Lambda(t-1) + \alpha(t)\alpha^T(t) + \beta(t)\beta^T(t)]$$

to get $V(t)$ and $D(t)$

- 10: Update $Q(t)$ and $\Lambda(t)$ [Equations(21)(22)]
- 11: Find the MC $\underline{q}(t)$ as $\hat{\vartheta}(t)$
- 12: Get $\hat{\theta}(t)$ from $\hat{\vartheta}(t)$
- 13: **end for**

C. Convergence Analysis

Convergence properties of the proposed algorithm are presented in the following theorem.

Theorem 1: For the system described by (1), $\hat{\theta}(t)$ converges to θ as $r \rightarrow \infty$ and $t \rightarrow \infty$ ($r \ll t$).

Proof: First, note that two facts hold:

- 1) $\widehat{\text{Var}}\{\omega(t)\} \rightarrow \text{Var}\{\omega(t)\}$, $\widehat{\text{Var}}\{v(t)\} \rightarrow \text{Var}\{v(t)\}$, and $\hat{\varepsilon} \rightarrow \varepsilon$ as $r \rightarrow \infty$;
- 2) $R(t) \rightarrow \text{Var}\{\psi(t)\}$ as $t \rightarrow \infty$.

Define

$$\begin{aligned} \psi_a(t) &= \begin{bmatrix} -\varepsilon y_o(t) & -\varepsilon y_o(t-1) & \cdots & -\varepsilon y_o(t-n_a) \\ u_o(t) & u_o(t-1) & \cdots & u_o(t-n_b) \end{bmatrix}^T, \\ \psi_b(t) &= \begin{bmatrix} -\lambda(t)\varepsilon y_o(t) & -\lambda(t-1)\varepsilon y_o(t-1) & \cdots \\ -\lambda(t-n_a)\varepsilon y_o(t-n_a) & \gamma(t)u_o(t) \\ \gamma(t-1)u_o(t-1) & \cdots & \gamma(t-n_b)u_o(t-n_b) \end{bmatrix}^T, \\ \eta(t) &= \begin{bmatrix} -\lambda(t)\varepsilon v(t) & -\lambda(t-1)\varepsilon v(t-1) & \cdots \\ -\lambda(t-n_a)\varepsilon v(t-n_a) & \gamma(t)\omega(t) \\ \gamma(t-1)\omega(t-1) & \cdots & \gamma(t-n_b)\omega(t-n_b) \end{bmatrix}^T. \end{aligned}$$

It is clear that

$$\psi_a^T(t)\vartheta = 0, \quad (23)$$

$$\psi(t) = \psi_b(t) + \eta(t) \quad (24)$$

From (23), we have

$$\vartheta^T \text{Var}\{\psi_a(t)\}\vartheta = 0. \quad (25)$$

Also note that

$$\text{Var}\{\psi_b(t)\} = p \text{Var}\{\psi_a(t)\}. \quad (26)$$

Let $\sigma^2 = \varepsilon^2 \text{Var}\{v(t)\} = \text{Var}\{\omega(t)\}$. Then

$$\text{Var}\{\eta(t)\} = p\sigma^2 I,$$

where I is an identity matrix of dimension $n_a + n_b + 2$. Therefore, as $r \rightarrow \infty$ and $t \rightarrow \infty$, $\hat{\vartheta}(t)$, which is equal to the MC of $\text{Var}\{\psi(t)\}$, is given by

$$\begin{aligned} \hat{\vartheta}^T(t)R(t)\hat{\vartheta}(t) &= \hat{\vartheta}^T(t)\text{Var}\{\psi(t)\}\hat{\vartheta}(t) \\ &= \hat{\vartheta}^T(t) [\text{Var}\{\psi_b(t)\} + p\sigma^2 I] \hat{\vartheta}(t) \\ &= p\hat{\vartheta}^T(t)\text{Var}\{\psi_a(t)\}\hat{\vartheta}(t) + p\sigma^2 \\ &\geq p\sigma^2. \end{aligned} \quad (27)$$

By comparing (25) and (27), the equality holds if and only if $\hat{\vartheta}(t) = \vartheta$. This completes the proof. ■

Remark 1: In [10], an efficient algorithm was developed; however, the availability probability p must be no less than $1/2$ in order to ensure consistency. In our approach, as in Theorem 1, no assumption is made about p . It follows that the proposed algorithm is convergent as long as $p > 0$ and N is large enough.

IV. SIMULATION STUDIES

Two numerical examples are given in this section to evaluate the performance of the proposed algorithm.

Example 1: Consider the following dynamic system:

$$y_o(t) = \frac{1 + 0.2z^{-1} + 0.5z^{-2}}{1 - 0.6z^{-1} + 0.8z^{-2}} u_o(t),$$

for which the normalized parameter vector is $\theta = [0.5513 \quad -0.3308 \quad 0.4411 \quad 0.5513 \quad 0.1103 \quad 0.2757]^T$. Here, $\{u_o(t)\}$ is taken as a white noise sequence with unit variance. $\{u_o(t)\}$ and $\{y_o(t)\}$ are corrupted respectively by white Gaussian noise sequences $\{w(t)\}$ and $\{v(t)\}$ with variances 0.1^2 . We assume that only 75% of both $\{u_o(t)\}$ and $\{y_o(t)\}$ are available, i.e., $p = 0.75$.

The proposed algorithm is implemented to estimate the parameters of the system. We select r as 500 to calculate $\hat{\varepsilon}$. Then the estimation procedures begin at time instant 501. Define the relative parameter estimation as

$$\begin{aligned} \delta\% &= \frac{(\hat{\theta} - \theta)^T(\hat{\theta} - \theta)}{\theta^T \theta} \times 100\% \\ &= [(\hat{\theta} - \theta)^T(\hat{\theta} - \theta)] \times 100\%. \end{aligned}$$

Intermediate parameter estimates and corresponding relative estimation errors are given in Table I. The parameter estimation errors $\delta\%$ versus t is shown in Fig. 2. From both it is seen that $\delta\%$ has a clearly decreasing tendency (in general). This shows the effectiveness of the proposed algorithm.

Example 2: In practice, it is of paramount importance to estimate parameters when there exists a large amount of missing data. Consider the same system as in Example 1, and suppose that $p = 0.2$ and $N = 200,000$, and take $r = 50,000$. As $p < 0.5$, the method in [10] cannot work in this case. By applying the approach in this paper, the parameter estimation errors $\delta\%$ versus t is illustrated in Fig. 3 and Table II. It exhibits good estimation performance, though the values of $\delta\%$ fluctuate at early time due to such a small p . It is worth mentioning that a large enough N is required for a small p .

V. CONCLUSION

This paper considers the problem of parameter estimation in the presence of a large percentage of missing input/output data. A recursive algorithm is proposed. In the approach, data missing is assumed to result from fictitious measurement noises. Meanwhile, MCA is known to be strongly robust to noises and thus very suitable in solving the proposed problem. Then an RMCA based estimation algorithm is presented, and its high effectiveness is illustrated by numerical simulation.

REFERENCES

- [1] W. Zhang, M.S. Branicky and S.M. Phillips, "Stability of networked control systems," *IEEE Control Systems Magazine*, vol. 21, no. 1, pp. 84-99, 2001.
- [2] Y. Shi and H. Fang, "Kalman filter based adaptive control for networked systems with unknown parameters and randomly missing outputs," *International Journal of Robust and Nonlinear Control – Special Issue Control with Limited Information*, in press, 2008.

TABLE I: Example 1: Intermediate parameter estimates and relative estimation errors

t	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{b}_0	\hat{b}_1	\hat{b}_2	$\delta\%$
504	0.209103	0.579698	0.351209	-0.0360828	0.643155	-0.286231	137.801
600	0.464202	-0.584608	0.396339	0.454604	-0.276859	0.0484676	53.3709
750	0.53269	-0.443675	0.459759	0.539379	-0.024394	0.128418	23.1074
1000	0.534574	-0.438544	0.452897	0.532313	-0.0326925	0.179913	20.493
2000	0.550077	-0.398198	0.458039	0.523997	0.0478876	0.228447	10.8166
4000	0.552948	-0.36443	0.45079	0.532258	0.0932737	0.257351	4.70761
θ	0.5513	-0.3308	0.4411	0.5513	0.1103	0.2757	

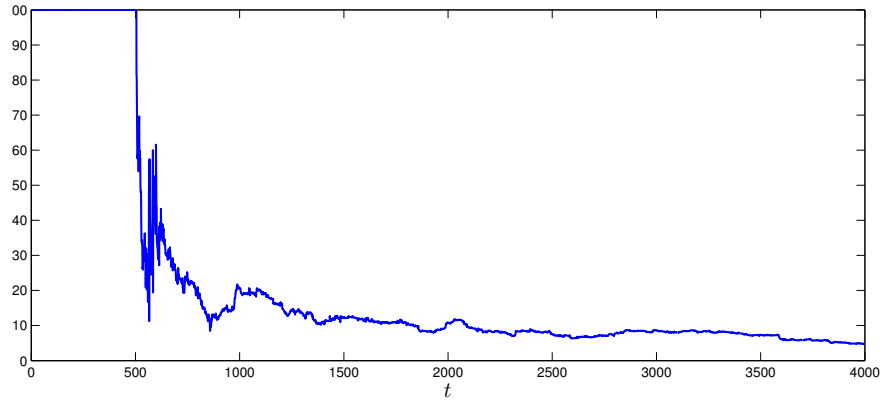


Fig. 2: Example 1: The relative parameter estimation errors versus time.

TABLE II: Example 2: Intermediate parameter estimates and relative estimation errors

t	\hat{a}_0	\hat{a}_1	\hat{a}_2	\hat{b}_0	\hat{b}_1	\hat{b}_2	$\delta\%$
50004	0.0752848	-0.43617	-0.211018	-0.769915	-0.407089	-0.0326895	166.472
60000	0.429417	-0.0530462	0.319979	0.557362	0.40428	0.486114	48.7273
75000	0.547209	-0.237188	0.430896	0.529615	0.243317	0.344872	17.8456
100000	0.57435	-0.386642	0.47786	0.480725	0.108709	0.222184	11.3357
150000	0.566088	-0.364806	0.472996	0.489296	0.101109	0.270373	7.97115
200000	0.556251	-0.339326	0.456729	0.515664	0.122585	0.293095	4.54963
θ	0.5513	-0.3308	0.4411	0.5513	0.1103	0.2757	

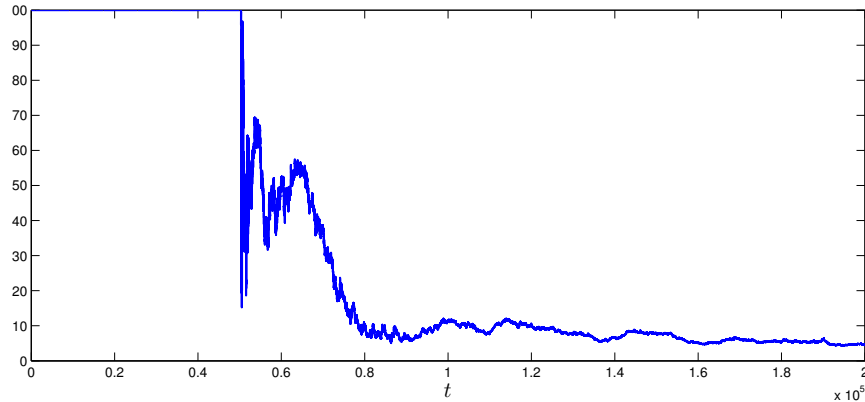


Fig. 3: Example 2: The relative parameter estimation errors versus time.

[3] Y. Shi and B. Yu, "Output feedback stabilization of networked control systems with random delays modeled by Markov chains," *IEEE Trans. on Automatic Control*, in press, 2009.

[4] 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.

- [5] R. Sanchis and P. Albertos, "Recursive identification under scarce measurements—convergence analysis," *Automatica*, vol. 38, no. 3, pp. 535-544, 2002.
- [6] A.J. Isaksson, "Identification of ARX-models subject to missing data," *IEEE Trans. Automat. Contr.*, vol. 38, no. 5, pp. 813-819, 1993.
- [7] F. Ding and T. Chen, "Identification of dual-rate systems based on finite impulse response models," *International Journal of Adaptive Control and Signal Processing*, vol. 18, no. 7, pp. 589-598, 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] R. Pintelon and J. Schoukens, "Frequency domain system identification with missing data," *IEEE Trans. Automat. Contr.*, vol. 45, no. 2, pp. 364-369, 2000.
- [10] J.M. Chen and B.S. Chen, "System parameter estimation with input/output noisy data and missing measurements," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1548-1558, 2000.
- [11] R. Wallin, A.J. Isaksson and L. Ljung, "An iterative method for identification of ARX models from incomplete data," *Proc. IEEE Conf. Decision & Control*, vol. 1, pp. 203-208, 2000.
- [12] B. Huang, "Process identification based on last principal component analysis," *Journal of Process Control*, vol. 11, pp. 19-33, 2001.
- [13] W. Li, H.H. Yue, S. Valle-Cervantes and S.J. Qin, "Recursive PCA for adaptive process monitoring," *Journal of Process Control*, vol. 10, pp. 471-486, 2000.