2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Stochastic Gradient Based Iterative Identification Algorithm for a Class of Dual-rate Wiener Systems

Jing Leng, Junpeng Li, Changchun Hua, Xinping Guan

Abstract— **Parameter estimation problem is considered for a class of dual-rate Wiener systems whose input-output data are measured by two different sampling rate. Firstly, a polynomial transformation technique is used to derive a mathematical model for such dual-rate Wiener systems. Then, directly based on the dual-rate sampled data, a dual-rate Wiener systems stochastic gradient algorithm (DRW-SG) is presented. In order to improve the algorithm convergence rate, a dual-rate Wiener systems stochastic gradient algorithm with a forgetting factor algorithm (DRW-FF-SG) is presented. For making full use of the forgetting factor, a dual-rate Wiener systems stochastic gradient algorithm with an increasing forgetting factor algorithm (DRW-IFF-SG) is presented which performs excellently. Finally, an example is provided to test and illustrate the proposed algorithms.**

I. Introduction

Wiener systems are block-oriented systems which are useful nonlinear dynamical systems [1]. They consist of interconnected linear dynamic systems and nonlinear static systems. Wiener systems identification has attracted much attention in the last decades [2-5]. However, most of the works are about single-rate Wiener systems.

Systems operating at different input and output sampling rates are called multirate systems [6-9] which can find many engineering applications, e.g., in digital signal processing [10], sensor networks [11], communications [12], process control [13] and so on. The parameter estimation problem is quite important in the analysis and design for multirate systems. In this paper, we focus on the identification of dual-rate Wiener systems.

Since the output $y(kq)$ is sampled at a slower rate than the input $u(k)$, the intersample outputs $\{y(kq+j), j = 1, 2, ..., q-1\}$ are missing. In this case, Shumway and Stoffer used an expectation maximization (EM) algorithm to handle the missing measurements for the linear state-space models [14]. These results have been extended to handle the nonlinear state-space models in [15] and [16]. The identification problems of ARX models with missing data is studied by Isaksson based on the Kalman filtering (fixed-interval smoothing) technique and maximum likelihood (ML) methods [17]. Sheng et al. [18] discussed model-based predictive control of multirate systems in the process control area. In the process identification literature, Li et al. [19] assumed that the system states were known, and they used them and the multirate input-output data to estimate the parameters of lifted state-space models for multirate systems. They attempted to extract fast single-rate models from the obtained lifted models, but the accuracy of the single-rate models was limited by that of the lifted models because the model conversion error was amplified greatly.

Based on the auxiliary model identification principle, Ding and Chen proposed a recursive least squares algorithm and a gradient-based recursive algorithm for the dual-rate output error type systems in [20] and [21], respectively. Then, the consistency of the SG algorithm for dual-rate sampled data systems was analyzed in [22]. In [23], Ding proposed a new hierarchical least squares algorithm for the dual-rate ARMAX systems. Recently, Chen et al. proposed a modified SG algorithm and a multi-innovation SG algorithm for the dual-rate Hammerstein system with preload nonlinearity in [24] and [25], respectively. On the other hand, multirate sampled-data systems were treated as missing data systems, and the expectation maximization (EM) algorithm was employed to estimate the parameters [26-28]. However, when too many data are missing, the EM algorithm results in poor parameter estimation accuracy.

In this paper, the parameter estimation problem is considered for a class of dual-rate Wiener systems whose static nonlinear block is expressed by the sum of known nonlinear (orthogonal or nonorthogonal) basis functions and unknown coefficients [29]. The polynomial transformation technique is introduced to derive a mathematical model for such dual-rate Wiener systems. Based on the mathematical model we use a dual-rate Wiener systems stochastic gradient algorithm (DRW-SG) to estimate the parameters directly using the dual-rate sampled data. In order to improve the convergence rate of the DRW-SG algorithm, a forgetting factor is introduced which is dual-rate Wiener systems stochastic gradient with a forgetting factor algorithm (DRW-FF-SG). Lastly, we analyze the affect of the forgetting factor on the algorithm, a dual-rate Wiener systems stochastic gradient with an increasing forgetting factor algorithm (DR-IFF-SG) is proposed which perform excellently.

The rest of the paper is organized as follows. In section 2, the identification problem formulation for dual-rate Wiener systems is described. Section 3 derive identification algorithms for dual-rate Wiener systems. Section 4 provides an illustrative example and compare the performance of the proposed algorithms. Finally, concluding remarks are given in section 5.

Jing Leng, Junpeng Li, Changchun Hua and Xinping Guan are with the Department of Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei, China (email:zicheyue.324@163.com).

This work was supported by the Major Program of the National Natural Science Foundation of China (No.61290322)

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

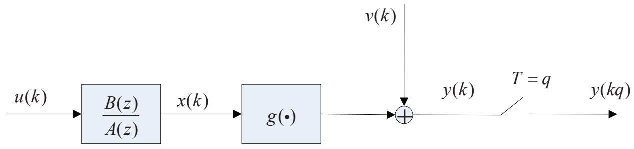

Fig. 1. The dual-rate sampled-data Wiener system.

II. Problem formulation

Consider the following dual-rate Wiener system which consists of a linear dynamic block followed by a static nonlinear subsystem, shown in Fig.1:

$$x(k) = \frac{B(z)}{A(z)}u(k) \tag{1}$$

$$y(k) = g(x(k)) + v(k) \tag{2}$$

Where $u(k)$ and $x(k)$ are the input and output of the linear dynamic block, respectively. $y(k)$ is the system output, and $y(kq)$ is the measurable system output whose sampling period is q times of $y(k)$. $v(k)$ is a white noise sequence with zero mean and variance σ^2, and $A(z)$ and $B(z)$ are polynomials as follows

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \tag{3}$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b} \tag{4}$$

where the parameters a_i and b_i are unknown, the orders n_a and n_b are assumed to be known, and z^{-1} is the unit backward shift operator, i.e., $z^{-1}y(k) = y(k-1)$.

The inner variables $x(k)$ is unmeasurable and $g(\cdot)$ is a static nonlinear function, including the piecewise nonlinearity, the monotonous (odd) nonlinearity, the polynomial nonlinearity and so on. Here, we assumed that the nonlinearity $y(k) = g(x(k))$ is the sum of the known nonlinear (orthogonal or nonorthogonal) basis functions $(g_1, g_2, \cdots, g_{n_\gamma})$ and unknown coefficients γ_i as follows.

$$\begin{aligned} g(x(k)) &= \gamma_1 g_1(x(k)) + \gamma_2 g_2(x(k)) + \cdots + \gamma_{n_\gamma} g_{n_\gamma}(x(k)) \\ &= \prod_{i=1}^{n_\gamma} \gamma_i g_i(x(k)) \end{aligned} \tag{5}$$

The following assumptions are made about the system:the linear dynamic system is asymptotically stable. The static nonlinear function can be expressed as a polynomial:

$$\begin{aligned} g(x(k)) &= \gamma_1 x(k) + \gamma_2 x(k)^2 + \cdots + \gamma_{n_\gamma} x(k)^{n_\gamma} \\ &= \prod_{i=1}^{n_\gamma} \gamma_i x(k)^i \end{aligned} \tag{6}$$

where the polynomial order n_γ is known. The inner output $x(k)$ of the linear dynamic system is unmeasurable. In order to get unique parameter estimate, we need to fix a coefficient of the nonlinear block. Here, we let the first entry of γ_i be unity, i.e., $\gamma_1 = 1$. Then, the static nonlinear function can be rewritten as

$$g(x(k)) = x(k) + \prod_{i=2}^{n_\gamma} \gamma_i x(k)^i \tag{7}$$

Formula (1) can be rewritten as

$$x(k) = [1 - A(z)x(k)] + B(z)u(k) \tag{8}$$

Substituting it into (2) for the separated $x(t)$, the system output is written in the form

$$y(k) = [1 - A(z)]x(k) + B(z)u(k) + \prod_{i=2}^{n_\gamma} \gamma_i x(k)^i + v(k) \tag{9}$$

However, notice that the available output in Fig.1 is $y(kq)$, that is, The outputs $y(kq - r)$ is missing when r is not a multiple of q. In this paper, we would make use of the polynomial transformation technique to obtain a new model that can directly use the dual-rate sampled data $\{u(k), y(kq)\}$. Then we can make use of the stochastic gradient algorithm and improved stochastic gradient algorithm to estimate the parameters.

III. The stochastic gradient algorithm and the improved stochastic gradient algorithm

Let us define the roots of $A(z)$ be $z_i (i = 1, 2, \cdots, n_a)$ to get

$$A(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) \cdots (1 - z_{n_a} z^{-1}) \tag{10}$$

Define the polynomials

$$\begin{aligned} G(z) &= \prod_{i=1}^{n_a} (1 + z_i z^{-1} + z_i^2 z^{-2} + \cdots + z_i^{q-1} z^{1-q}) \\ &= 1 + g_1 z^{-1} + g_2 z^{-2} + \cdots + g_{n_g} z^{-n_g}, \quad n_g = (q-1)n_a \end{aligned} \tag{11}$$

$$\begin{aligned} \alpha(z) &= G(z)A(z) \\ &= 1 + \alpha_1 z^{-q} + \alpha_2 z^{-2q} + \cdots + \alpha_{n_a} z^{-n_a q} \end{aligned} \tag{12}$$

$$\begin{aligned} \beta(z) &= G(z)B(z) \\ &= \beta_1 z^{-1} + \beta_2 z^{-2} + \cdots + \beta_m z^{-m}, \quad m = n_g + n_b \end{aligned} \tag{13}$$

Multiplying both sides of (1) by $G(z)$ gives

$$x(k) = \frac{\beta(z)}{\alpha(z)}u(k) \tag{14}$$

The above formula can be rewritten as

$$x(k) = [1 - \alpha(z)x(k)] + \beta(z)u(k) \tag{15}$$

Substituting it into (2) for the separated $x(t)$, the system output is written in the form

$$y(k) = [1 - \alpha(z)]x(k) + \beta(z)u(k) + \prod_{i=2}^{n_\gamma} \gamma_i x(k)^i + v(k) \tag{16}$$

Define the parameter vector $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ as

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_{n_a}]^T \in R^{n_a} \tag{17}$$

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_m]^T \in R^m \tag{18}$$

$$\boldsymbol{\gamma} = [\gamma_2, \gamma_3, \cdots, \gamma_m]^T \in R^{n_\gamma - 1} \tag{19}$$

Define the information vector $\boldsymbol{\varphi}(k)$, $\boldsymbol{\phi}(k)$ and $\boldsymbol{\psi}(k)$ as

$$\boldsymbol{\varphi}(k) = [-x(k-q), -x(k-2q), \cdots, -x(k-n_a q)]^T \tag{20}$$

$$\boldsymbol{\phi}(k) = [-u(k-1), -u(k-2), \cdots, -u(k-m)]^T \tag{21}$$

$$\boldsymbol{\psi}(k) = [x^2(k), x^3(k), \cdots, x^{n_\gamma}(k)]^T \tag{22}$$

From (16)-(22), we have

$$y(k) = \boldsymbol{\varphi}^T(k)\boldsymbol{\alpha} + \boldsymbol{\phi}^T(k)\boldsymbol{\beta} + \boldsymbol{\psi}^T(k)\boldsymbol{\gamma} + v(k) \tag{23}$$

Define the parameter vector θ and the information vector Ψ as follows

$$\boldsymbol{\theta} = [\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T]^T \in R^{n_a+m+n_\gamma-1} \tag{24}$$

$$\boldsymbol{\Psi}(k) = [\boldsymbol{\varphi}^T(k), \boldsymbol{\phi}^T(k), \boldsymbol{\psi}^T(k)]^T \in R^{n_a+m+n_\gamma-1} \tag{25}$$

Then, formula (23) can be written as

$$y(k) = \boldsymbol{\Psi}^T(k)\boldsymbol{\theta} + v(k) \tag{26}$$

Replaceing k with kq gives

$$y(kq) = \boldsymbol{\Psi}^T(kq)\boldsymbol{\theta} + v(kq) \tag{27}$$

Notice that $\alpha(z)$ and $\beta(z)$ are polynomials in z^{-q} and z^{-1}, respectively. Thus the above model can be identified directly by using the dual-rate sample data $\{u(k), y(kq)\}$. However, the difficulty of identification is that $\boldsymbol{\varphi}(kq)$ contains the unknown variables $x(kq-iq)$. The solution is to replace $x(kq-iq)$ with its estimate $\hat{x}(kq-iq)$ and to be computed as follows.

$$\hat{x}(kq-iq) = \hat{\boldsymbol{\varphi}}^T(kq-iq)\hat{\boldsymbol{\alpha}}(kq-iq) + \boldsymbol{\phi}^T(kq-iq)\hat{\boldsymbol{\beta}} \tag{28}$$

$\hat{\boldsymbol{\varphi}}(kq)$ is the estimate of $\boldsymbol{\varphi}(kq)$ as follows

$$\hat{\boldsymbol{\varphi}}(kq) = [-\hat{x}(kq-q), -\hat{x}(kq-2q), \cdots, -\hat{x}(kq-n_a q)]^T \tag{29}$$

$\hat{\boldsymbol{\psi}}(kq)$ is the estimate of $\boldsymbol{\psi}(kq)$ as follows

$$\hat{\boldsymbol{\psi}}(kq) = [\hat{x}^2(kq), \hat{x}^3(kq), \cdots, \hat{x}^{n_\gamma}(kq)]^T \tag{30}$$

Based on the square criterion we have

$$J(\hat{\boldsymbol{\theta}}) = \|y(kq) - \boldsymbol{\Psi}^T(kq)\boldsymbol{\theta}\|^2 \tag{31}$$

In order to minimizing the above cost function, we can use the stochastic gradient search algorithm [33] as follows

$$\hat{\boldsymbol{\theta}}(kq) = \hat{\boldsymbol{\theta}}(kq-q) + \frac{1}{r(kq)}\hat{\boldsymbol{\Psi}}(kq)[y(kq) - \hat{\boldsymbol{\Psi}}^T(kq)\hat{\boldsymbol{\theta}}(kq-q)] \tag{32}$$

Eq.(32) can be rewritten as

$$\hat{\boldsymbol{\theta}}(kq) = [\boldsymbol{I} - \frac{1}{r(kq)}\hat{\boldsymbol{\Psi}}(kq)\hat{\boldsymbol{\Psi}}^T(kq)]\hat{\boldsymbol{\theta}}(kq-q) + \frac{1}{r(kq)}\hat{\boldsymbol{\Psi}}(kq)y(kq) \tag{33}$$

Note that Eq.(33) is equivalent to one discrete-time system with the state $\hat{\boldsymbol{\theta}}$. In order to guarantee the convergence of the parameters $\hat{\boldsymbol{\theta}}$, the symmetric matrices $[\boldsymbol{I} - \frac{1}{r(kq)}\hat{\boldsymbol{\Psi}}(kq)\hat{\boldsymbol{\Psi}}^T(kq)]$ need to have all eigenvalues inside the unit circle. One conservative choice is to have the convergence factors to satisfy

$$0 < \frac{1}{r(kq)} \leq \frac{2}{\mu_{max}[\hat{\boldsymbol{\Psi}}(kq)\hat{\boldsymbol{\Psi}}^T(kq)]} \tag{34}$$

Here, $\mu_{max}[\hat{\boldsymbol{\Psi}}(kq)\hat{\boldsymbol{\Psi}}^T(kq)]$ is the maximum eigenvalues of the matrix $\hat{\boldsymbol{\Psi}}(kq)\hat{\boldsymbol{\Psi}}^T(kq)$. The convergence factor $\frac{1}{r(kq)}$ is updated as follows

$$r(kq) = r(kq-q) + \|\hat{\boldsymbol{\Psi}}(kq)\|^2, \quad r(0) = 1 \tag{35}$$

Therefore, the dual-rate Wiener systems stochastic gradient algorithm (DRW-SG) can be summarized as follows:

$$\hat{\boldsymbol{\theta}}(kq) = \hat{\boldsymbol{\theta}}(kq-q) + \frac{1}{r(kq)}\hat{\boldsymbol{\Psi}}(kq)[y(kq) - \hat{\boldsymbol{\Psi}}^T(kq)\hat{\boldsymbol{\theta}}(kq-q)] \tag{36}$$

$$r(kq) = r(kq-q) + \|\hat{\boldsymbol{\psi}}(kq)\|^2, \quad r(0) = 1 \tag{37}$$

$$\hat{\boldsymbol{\Psi}}(kq) = [\hat{\boldsymbol{\varphi}}(kq), \boldsymbol{\phi}(kq), \hat{\boldsymbol{\psi}}(kq)]^T \tag{38}$$

$$\hat{x}(kq-iq) = \hat{\boldsymbol{\varphi}}^T(kq-iq)\hat{\boldsymbol{\alpha}}(kq-iq) + \boldsymbol{\phi}^T(kq-iq)\hat{\boldsymbol{\beta}} \tag{39}$$

$$\hat{\boldsymbol{\varphi}}(kq) = [-\hat{x}(kq-q), -\hat{x}(kq-2q), \cdots, -\hat{x}(kq-n_a q)]^T \tag{40}$$

$$\boldsymbol{\phi}(k) = [-u(k-1), -u(k-2), \cdots, -u(k-m)]^T \tag{41}$$

$$\hat{\boldsymbol{\psi}}(kq) = [\hat{x}^2(kq), \hat{x}^3(kq), \cdots, \hat{x}^{n_\gamma}(kq)]^T \tag{42}$$

$$\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\gamma}}^T]^T \tag{43}$$

$$\hat{\boldsymbol{\alpha}} = [\hat{\alpha_1}, \hat{\alpha_2}, \cdots, \hat{\alpha_{n_a}}]^T \tag{44}$$

$$\hat{\boldsymbol{\beta}} = [\hat{\beta_1}, \hat{\beta_2}, \cdots, \hat{\beta_m}]^T \tag{45}$$

$$\hat{\boldsymbol{\gamma}} = [\hat{\gamma_2}, \hat{\gamma_3}, \cdots, \hat{\gamma_m}]^T \tag{46}$$

Just as verified in [29,30], the stochastic gradient algorithm has low convergency rate. In order to improve the convergence rate of the DRW-SG algorithm, we introduce a forgetting factor λ in the DRW-SG algorithm to get a dual-rate Wiener systems stochastic gradient algorithm with a forgetting factor, which is abbreviated as DRW-FF-SG.

$$r(kq) = \lambda r(kq-q) + \|\hat{\boldsymbol{\Psi}}(kq)\|^2, \quad 0 < \lambda \leq 1; \;\; r(0) = 1 \tag{47}$$

In fact, when the forgetting factor λ is small, the algorithm's convergence rate is fast. However, the algorithm's convergence stationarity is bad. When the forgetting factor λ is large, the algorithm shows the converse performance. During the iterative process of the algorithm, we hope that the convergence rate of the algorithm is fast at the beginning. With the increasing of the number of iterations the algorithm strengthens the convergence stationarity. Base on the above idea we proposed another improved dual-rate Wiener systems stochastic gradient algorithm with an increasing forgetting factor, which is abbreviated as DRW-IFF-SG.

$$r(kq) = \lambda(t)r(kq-q) + \|\hat{\mathbf{\Psi}}(kq)\|^2, \quad r(0) = 1 \tag{48}$$

The $\lambda(t)$ is define as follows

$$\lambda(t) = \lambda_{min} + (\lambda_{max} - \lambda_{min})(\frac{t}{t_{max}})^{\omega} \tag{49}$$

where λ_{max} and λ_{min} are the maximum and minimum values of λ. t_{max} is the maximum iterative numbers, and t is the present iterative numbers. ω is the increasing adjustment factor.

Remark 1. The DRW-IFF-SG algorithm fully make use of the forgetting factor's effect. At the beginning, the convergence rate of the algorithm is fast at the cost of the poor convergence stationarity of the algorithm. Finally, the algorithm strengthens the convergence stationarity at the cost of the slow convergence rate. This draws on the idea of swarm intelligent optimization that explore first, and then exploit. The increasing adjustment factor ω can control the balance of them. The DRW-IFF-SG algorithm performs excellently.

IV. Example

Consider the following dual-rate system model:

$$y(k) = \frac{B(z)}{A(z)}x(k) + v(k)$$

$$y(k) = g(x(k)) + v(k)$$

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} = 1.00 - 0.80z^{-1} + 0.16z^{-2}$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 0.50z^{-1} + 0.40z^{-2}$$

$$\begin{aligned} g(x(k)) =& \gamma_1 x(k) + \gamma_2 x(k) + \gamma_3 x(k) = x(k) + 0.40x^2(k) \\ &+ 0.25x^3(k) \end{aligned}$$

Here, we take $q = 2$ and $G(z) = 1 + 0.80z^{-1} + 0.16z^{-2}$, then have

$$\alpha(z) = G(z)A(z) = 1.00 - 0.32z^{-2} + 0.0256z^{-4}$$

$$\begin{aligned} \beta(z) =& G(z)B(z) = 0.50z^{-1} + 0.80z^{-2} + 0.40z^{-3} \\ &+ 0.064z^{-4} \end{aligned}$$

$$\boldsymbol{\theta} = [-0.32, 0.0256, 0.50, 0.80, 0.40, 0.064, 0.40, 0.25]^T$$

The input $\{u(k)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance, and $\{v(k)\}$ as a white noise sequence with zero mean and variance $\sigma^2 = 0.1^2$, the corresponding noise-to-signal ratio is $\delta_{ns} = 7.66\%$. To quantify the estimation accuracy, define the estimation error as $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$.

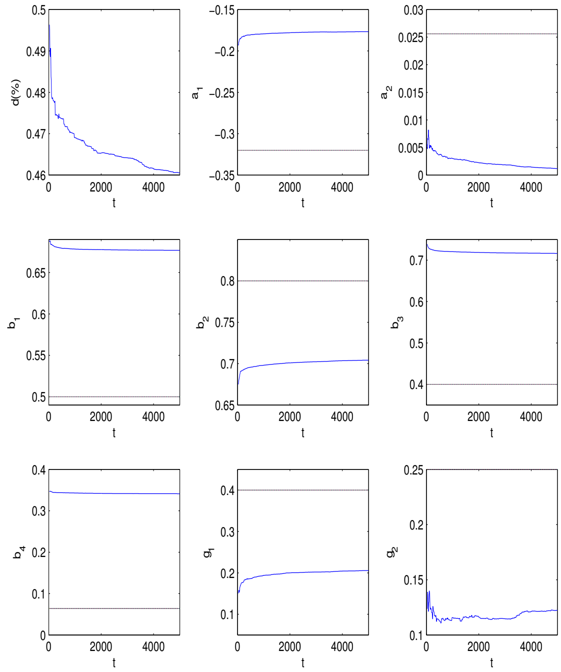

Fig. 2. The estimated parameters and the estimation error δ versus k using DRW-FF-SG algorithm with the forgetting factors $\lambda = 1$.

The estimated parameters and the estimation error δ versus k using DRW-FF-SG algorithm with the different forgetting factors λ are illustrated in Fig.3, Fig.4 and Fig.5. It is apparent that the DR-FF-SG algorithm with $\lambda = 1$ is equivalent to DRW-SG algorithm. To clearly compare the performance of DRW-FF-SG algorithm with the different forgetting factors λ, the estimated parameter values and the estimation error values δ versus k are illustrated in Table.1. From Fig.3-Fig.5 and Table.1, we can know that the convergence speed is very slow when the forgetting factorss $\lambda = 1$ which is equal to DRW-SG algorithm. In fact, the estimation error of DRW-SG algorithm is lager than 40% at the end of the iteration which is very bad. The convergence speed is fast when the forgetting factor $\lambda = 0.8$, and the estimation error is equal to 1.9818% at the end of the iteration. When the forgetting factor $\lambda = 0.6$ the convergence speed is faster than that with $\lambda = 0.8$. However, the convergence stationarity is worse. Due to the large trembling, the the estimation error is lager which is equal to 5.6509% at the end of the iteration.

Remark 2. The smaller the forgetting factor λ is, the

TABLE I

THE ESTIMATED PARAMETER VALUES AND THE ESTIMATION ERROR VALUES δ VERSUS k USING DRW-FF-SG ALGORITHM WITH THE DIFFERENT FORGETTING FACTORS λ.

λ	k	α_1	α_2	β_1	β_2	β_3	β_4	λ_1	λ_2	δ
$\lambda = 1$	100	-0.18433	0.0062062	0.68378	0.68819	0.72806	0.34577	0.1643	0.12446	48.415
	200	-0.18265	0.0049324	0.68183	0.69217	0.72557	0.34437	0.17735	0.12036	47.78
	500	-0.18045	0.0037748	0.67931	0.69567	0.72204	0.34368	0.18661	0.11166	47.371
	1000	-0.17937	0.0030277	0.67841	0.69805	0.72055	0.34302	0.1941	0.11563	46.907
	2000	-0.17803	0.0022353	0.67767	0.70093	0.71861	0.34204	0.20012	0.11614	46.536
	3000	-0.17716	0.0018697	0.67728	0.70233	0.7175	0.34153	0.20203	0.11471	46.418
	4000	-0.17704	0.0014306	0.67704	0.70357	0.71692	0.34139	0.20438	0.12144	46.15
	5000	-0.17676	0.0011767	0.67686	0.70421	0.71641	0.34119	0.20559	0.12231	46.054
$\lambda = 0.8$	100	-0.10917	-0.03267	0.62585	0.83007	0.61729	0.32254	0.26948	0.087492	40.122
	200	-0.10834	-0.071234	0.56653	0.89762	0.58676	0.27932	0.31648	0.15992	34.527
	500	-0.15705	-0.06482	0.51922	0.88997	0.55385	0.2436	0.31072	0.14754	29.194
	1000	-0.16803	-0.040232	0.52691	0.84337	0.48968	0.19876	0.36839	0.22024	20.525
	2000	-0.25869	-0.011541	0.51629	0.81354	0.44601	0.1055	0.37597	0.22386	8.8015
	3000	-0.29534	0.013545	0.50124	0.78639	0.4218	0.074949	0.41013	0.25568	3.4797
	4000	-0.31676	0.030161	0.47991	0.79688	0.3852	0.082695	0.41474	0.26192	3.154
	5000	-0.33249	0.022812	0.50374	0.80145	0.39496	0.054418	0.38445	0.25172	1.9818
$\lambda = 0.6$	100	-0.10352	-0.077461	0.53462	0.87468	0.57295	0.26343	0.30065	0.11297	34.329
	200	-0.13631	-0.061057	0.54783	0.91576	0.54438	0.20565	0.33396	0.1931	27.646
	500	-0.2325	-0.02293	0.48697	0.84078	0.48371	0.16162	0.35613	0.19612	15.515
	1000	-0.26579	0.015847	0.49233	0.78684	0.42177	0.093634	0.41276	0.27789	6.3473
	2000	-0.30428	0.011746	0.5206	0.80763	0.41789	0.046859	0.37777	0.22581	4.3537
	3000	-0.31038	0.014625	0.49617	0.78637	0.4274	0.060843	0.39078	0.23874	3.1691
	4000	-0.32888	0.044999	0.46123	0.81316	0.38113	0.077792	0.42177	0.26042	4.861
	5000	-0.33631	0.021131	0.51086	0.81859	0.40836	0.046342	0.34893	0.22419	5.6509
	True values	-0.32	0.0256	0.5	0.8	0.4	0.064	0.4	0.25	0

TABLE II

THE ESTIMATED PARAMETER VALUES AND THE ESTIMATION ERROR VALUES δ VERSUS k USING DRW-IFF-SG ALGORITHM WITH THE DIFFERENT INCREASING ADJUSTMENT FACTOR ω.

ω	k	α_1	α_2	β_1	β_2	β_3	β_4	λ_1	λ_2	δ
$\omega = 0.5$	100	-0.10112	-0.071888	0.54862	0.87458	0.57688	0.27992	0.29501	0.1059	35.619
	200	-0.12071	-0.07217	0.5487	0.9168	0.55073	0.22837	0.33133	0.19042	29.894
	500	-0.20064	-0.041639	0.50137	0.86497	0.51111	0.19464	0.33449	0.17232	21.323
	1000	-0.21566	-0.014923	0.51178	0.81822	0.45212	0.14934	0.39278	0.2481	12.927
	2000	-0.27919	0.0021706	0.50845	0.81012	0.43145	0.09174	0.38577	0.23208	5.8179
	3000	-0.29328	0.014258	0.50143	0.79305	0.41467	0.080479	0.40995	0.25774	3.3393
	4000	-0.29957	0.015067	0.49717	0.79418	0.40094	0.083194	0.40716	0.25768	2.7584
	5000	-0.3005	0.012916	0.49918	0.79428	0.40119	0.078997	0.40966	0.2568	2.6108
$\omega = 0.8$	100	-0.1029	-0.076164	0.53795	0.87487	0.57388	0.26765	0.29929	0.11108	34.655
	200	-0.131	-0.065149	0.54786	0.91612	0.54576	0.2131	0.33358	0.19264	28.354
	500	-0.21943	-0.030749	0.49372	0.85188	0.49386	0.17462	0.34713	0.18579	17.856
	1000	-0.24282	0.0015684	0.50025	0.80307	0.43439	0.11986	0.40653	0.26558	8.9826
	2000	-0.29906	0.012278	0.51094	0.80986	0.41894	0.068343	0.38626	0.23288	3.5022
	3000	-0.311	0.023001	0.50116	0.78738	0.40742	0.063652	0.41573	0.26135	2.22
	4000	-0.31605	0.023841	0.49613	0.79459	0.38967	0.071124	0.40853	0.25809	1.6125
	5000	-0.31664	0.020269	0.49896	0.79365	0.3923	0.066145	0.40982	0.25838	1.5029
$\omega = 1.0$	100	-0.1033	-0.076923	0.53596	0.87478	0.57331	0.26517	0.3001	0.11218	34.461
	200	-0.13374	-0.063102	0.54779	0.91591	0.54486	0.20928	0.33388	0.19296	27.979
	500	-0.22505	-0.027445	0.49118	0.84749	0.48917	0.16888	0.35101	0.19007	16.836
	1000	-0.25137	0.0066646	0.4969	0.79802	0.42928	0.11041	0.40975	0.27038	7.897
	2000	-0.30299	0.01384	0.51295	0.80982	0.41653	0.061681	0.38557	0.23222	3.2917
	3000	-0.31459	0.024328	0.5011	0.78575	0.40722	0.05978	0.41625	0.26117	2.2427
	4000	-0.31951	0.026107	0.49525	0.79495	0.38664	0.069372	0.40908	0.25847	1.7244
	5000	-0.32019	0.021757	0.49906	0.79347	0.39028	0.063842	0.40921	0.25864	1.5044
$\omega = 1.5$	100	-0.10351	-0.077396	0.53477	0.87469	0.57298	0.26363	0.30059	0.11288	34.343
	200	-0.13588	-0.061421	0.54781	0.91577	0.5444	0.20627	0.33398	0.1931	27.699
	500	-0.23056	-0.024138	0.48828	0.84274	0.48497	0.16344	0.35479	0.19445	15.854
	1000	-0.26079	0.012457	0.49368	0.79154	0.42406	0.099694	0.41231	0.27534	6.8304
	2000	-0.3053	0.013885	0.51648	0.80918	0.41554	0.054174	0.38355	0.23032	3.5154
	3000	-0.31662	0.02385	0.50095	0.78431	0.41007	0.056645	0.4134	0.25791	2.1859
	4000	-0.32279	0.0294	0.49177	0.79509	0.38312	0.069925	0.41052	0.26003	2.1629
	5000	-0.32387	0.023384	0.49916	0.79297	0.38778	0.06266	0.40739	0.25899	1.6075
$\omega = 3$	100	-0.10352	-0.07746	0.53462	0.87468	0.57295	0.26343	0.30065	0.11297	34.329
	200	-0.13631	-0.061059	0.54783	0.91576	0.54438	0.20565	0.33396	0.1931	27.646
	500	-0.23246	-0.022958	0.48701	0.84084	0.48373	0.16166	0.3561	0.19608	15.522
	1000	-0.2655	0.015625	0.4924	0.78721	0.42186	0.094017	0.4128	0.27774	6.3727
	2000	-0.30469	0.012288	0.51988	0.80799	0.41709	0.048439	0.37945	0.22707	4.1294
	3000	-0.31419	0.019184	0.49935	0.78424	0.4184	0.057449	0.40207	0.24831	2.272
	4000	-0.32585	0.034968	0.47924	0.79747	0.38043	0.075466	0.41421	0.26204	3.2088
	5000	-0.32869	0.02567	0.49991	0.7934	0.3861	0.061654	0.40183	0.25736	1.6491
	True values	-0.32	0.0256	0.5	0.8	0.4	0.064	0.4	0.25	0

faster the convergence speed is. But the convergence tremble is larger. On the contrary, the larger the forgetting factor λ is, the slower the convergence speed is. But the convergence stationarity is greater.

The estimated parameters and the estimation error δ versus k using DRW-IFF-SG algorithm with the different increasing adjustment factor ω are illustrated in Fig.6-Fig.10. To clearly compare the performance of DRW-DFF-SG algorithm with the different increasing adjustment factors λ, the estimated parameter values and the estimation error values δ versus k are illustrated in Table.2. Fig.2 shows the curvature of λ for the different increasing adjustment factor ($\omega = 0.5, 0.8, 1, 1.5, 3$). From Fig.6-Fig.10 and Table.2, we can know that the DRW-IFF-SG algorithm perform excellently. The increasing adjustment factor ω is used to adjust the increasing speed of the forgetting factor. When $\omega = 3$, the increasing speed of λ is small at the beginning stage, and is large at the final stage. Therefore, the DRW-IFF-SG algorithm's convergence rate is very fast at the beginning stage, and keep steady transitorily at the final stage. When $\omega = 0.5$, the increasing speed of λ is large at the beginning stage,

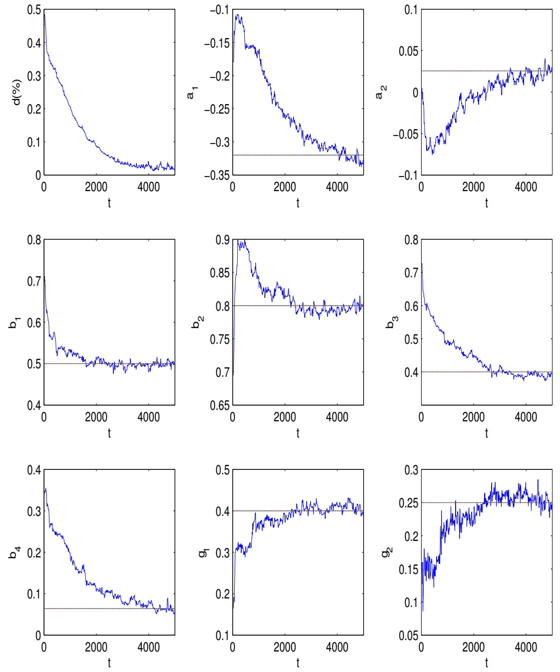

Fig. 3. The estimated parameters and the estimation error δ versus k using DRW-FF-SG algorithm with the forgetting factors $\lambda = 0.8$.

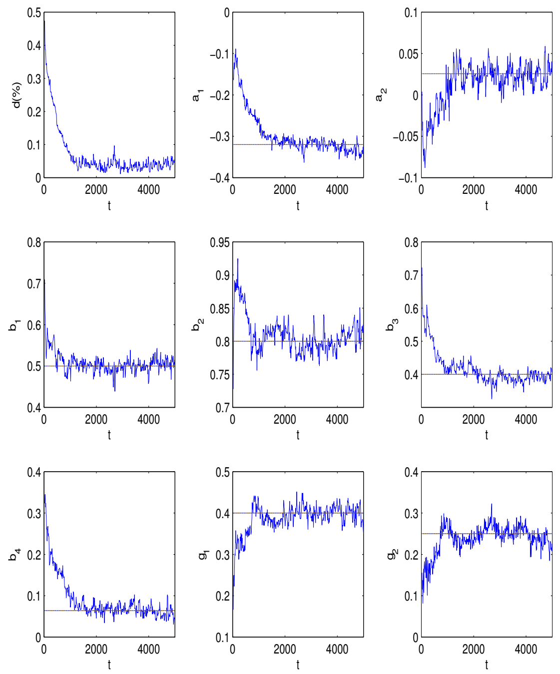

Fig. 4. The estimated parameters and the estimation error δ versus k using DRW-FF-SG algorithm with the forgetting factors $\lambda = 0.6$.

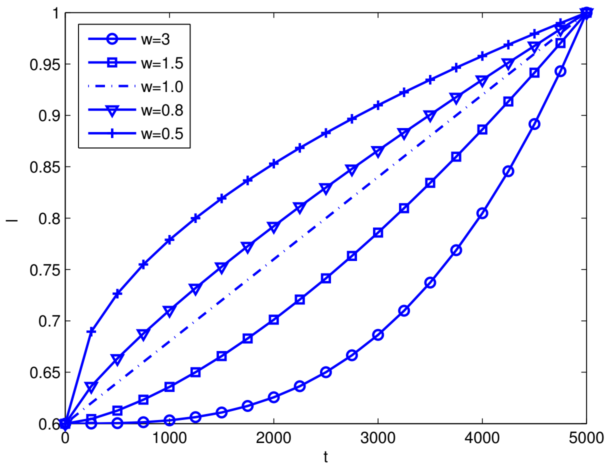

Fig. 5. The increasing forgetting factor λ versus k for the different increasing adjustment factor ω.

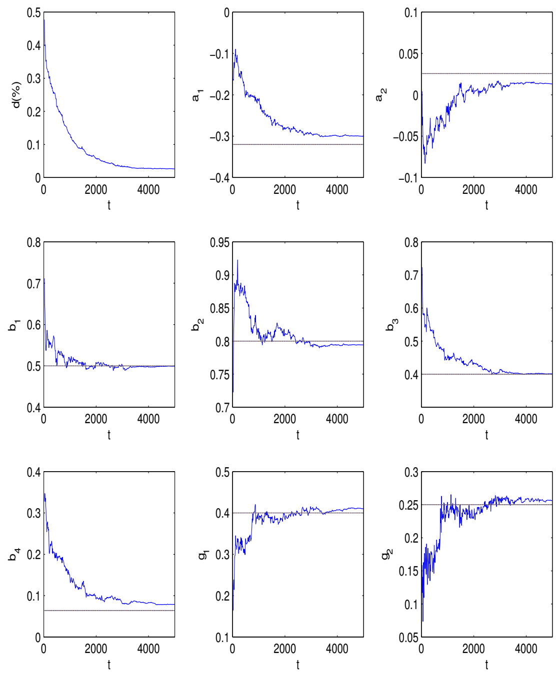

Fig. 6. The estimated parameters and the estimation error δ versus k using DRW-IFF-SG algorithm with the increasing adjustment factor $\omega = 0.5$.

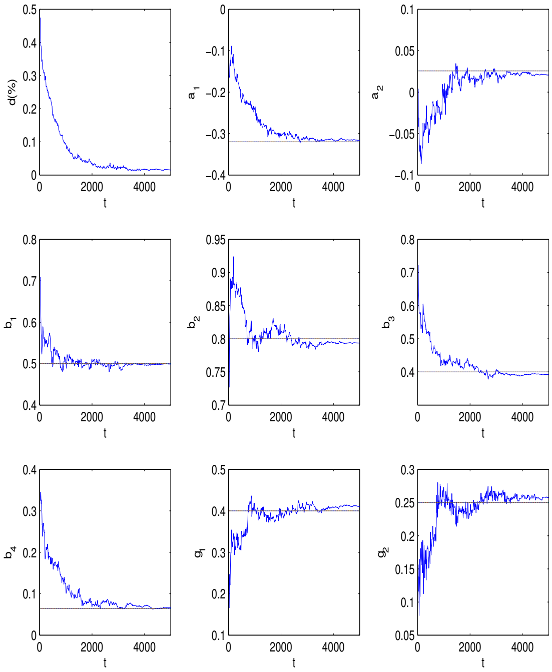

Fig. 7. The estimated parameters and the estimation error δ versus k using DRW-IFF-SG algorithm with the increasing adjustment factor $\omega = 0.8$.

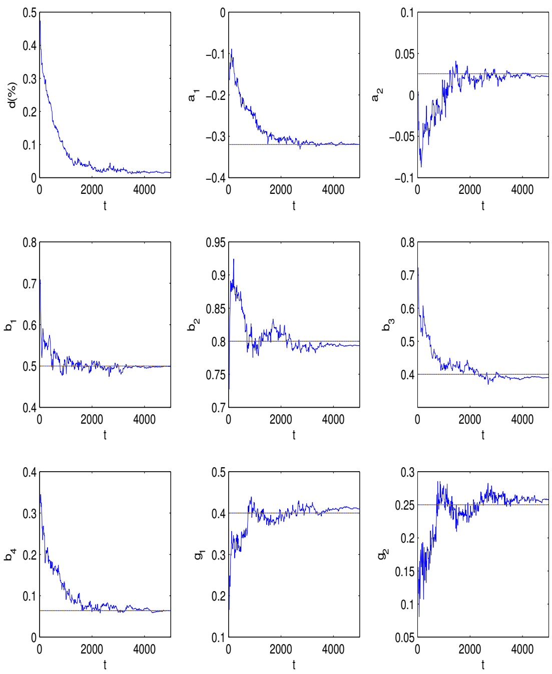

Fig. 8. The estimated parameters and the estimation error δ versus k using DRW-IFF-SG algorithm with the increasing adjustment factor $\omega = 1$.

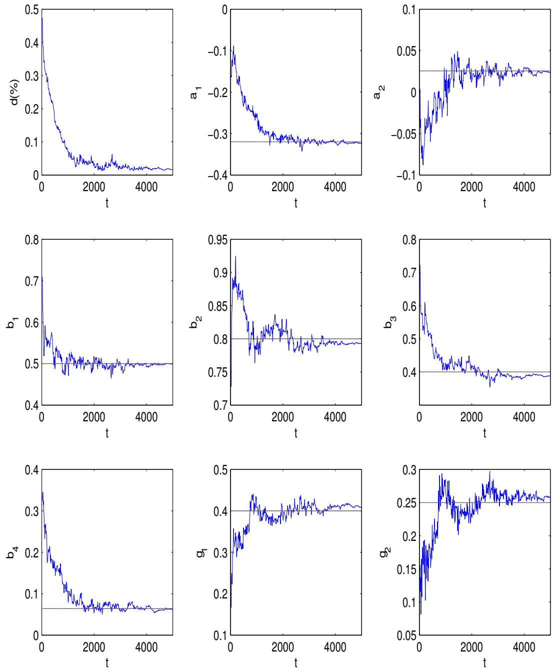

Fig. 9. The estimated parameters and the estimation error δ versus k using DRW-IFF-SG algorithm with the increasing adjustment factor $\omega = 1.5$.

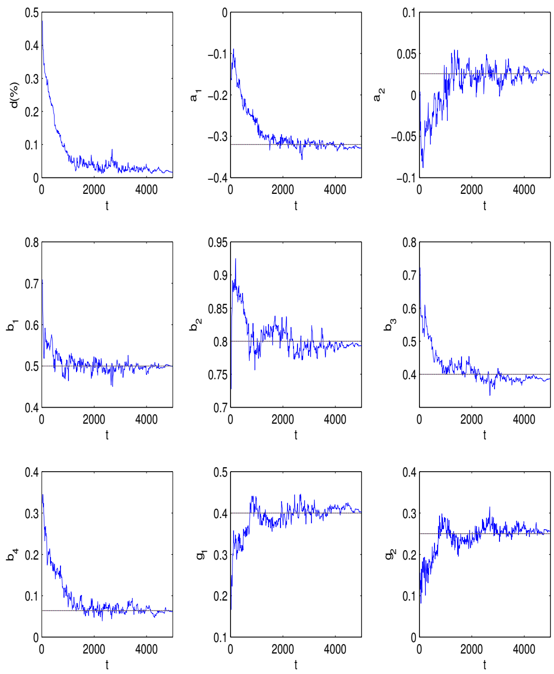

Fig. 10. The estimated parameters and the estimation error δ versus k using DRW-IFF-SG algorithm with the increasing adjustment factor $\omega = 3$.

and is small at the final stage. Therefore, the DRW-IFF-SG algorithm's convergence rate is fast in a short period of time at the beginning stage, and keep steady for a long time at the final stage. In fact, The increasing adjustment factor ω can control the balance of the DRW-IFF-SG algorithm's convergence rate and convergence stationarity. From the simulations, we can know that the DRW-IFF-SG algorithm performs the best with $\omega = 0.8$.

V. Conclusions

In this paper, a dual-rate Wiener systems stochastic gradient algorithm (DRW-SG) is presented. In order to improve the algorithm's convergence rate, a dual-rate Wiener systems stochastic gradient algorithm with a forgetting factor algorithm (DRW-FF-SG) is presented. The simulation shows the forgetting factor's impact on the performance of the algorithm. In order to further improve the performance of the algorithm, a dual-rate Wiener systems stochastic gradient algorithm with an increasing forgetting factor algorithm (DRW-IFF-SG) is presented which performs excellently. The numerical simulation confirms the established theorem.

References

[1] N. Wiener, Nonlinear problems in random theory, The MIT Press, USA, 1966.

[2] S. Nakamori, A. Hermoso-Carazo, J. Linares-Prez, Design of RLS Wiener estimators from randomly delayed observations in linear discrete-time stochastic systems, Applied Mathematics and Computation 217 (8) (2010) 3801-3815.

[3] K. Pelckmans, MINLIP for the identification of monotone Wiener systems, Automatica 47 (10) (2011) 2298-2305.

[4] M. Sznaier, W.J. Ma, O.I. Camps, H. Lim, Risk Adjusted Set Membership Identification of Wiener Systems, IEEE Transactions on Automatic Control 54 (5) (2009) 1147-1152.

[5] B.Q. Mu, H.F. Chen, Recursive Identification of MIMO Wiener Systems, IEEE Transactions on Automatic Control 58 (3) (2013) 802-808.

[6] J.H. Ahrens, X. Tan, H.K. Khalil, Multirate sampled-data output feedback control with application to smart material actuated systems, IEEE Transation on Automatic Control 54 (11) (2009) 2518-2529.

[7] J.S.H. Tsai, C.M. Chen, L.S. Shieh, Modelling of multirate feedback systems using uniform-rate models, Applied Mathematical Modelling 17 (1) (1993) 2-14.

[8] S. Roy, H.B. Yan, Blind channel estimation in multi-rate CDMA systems, IEEE Transactions on Communications 50 (6) (2002) 995-1004.

[9] S. Roshany-Yamchi, M. Cychowski, R.R. Negenborn, B.De Schutter, Kalman Filter-Based Distributed Predictive Control of Large-Scale Multi-Rate Systems: Application to Power Networks, IEEE Transactions on Control Systems Technology 21 (1) (2013) 27-39.

[10] S. Nakamori, A. Hermoso-Carazo, J. Linares-Prez, Suboptimal estimation of signals from uncertain observations using approximations of mixtures, Digital Signal Process 17 (1) (2007) 4-16.

[11] D.L. Xie, W.P. Wei, Y. Wang, H.S. Zhu, Tradeoff Between Throughput and Energy Consumption in Multirate Wireless Sensor Networks, IEEE Sensors Journal 13 (10) (2013) 3667-3676.

[12] Y. Shi, F. Ding, T. Chen, Multirate crosstalk identification in xDSL systems, IEEE Transaction on Communications 54 (10) (2006) 1878-1886.

[13] D. Li, S.L. Shah, T. Che, Analysis of dual-rate inferential control system, Automatica 38 (6) (2002) 1053C1059.

[14] R.H. Shumway, D.S. Stoffer, An approach to time series smoothing and forecasting using the em algorithm, Journal of Time Series Analysis 3 (4) (1982) 253-264.

[15] R.B. Gopaluni, identification of nonlinear processes with known structure under missing observations, in: 17th IFAC World Congress, Jul. 6-11, Seoul, Korea, 2008, pp.6478-6483

[16] R.B. Gopaluni, A particle filter approach to identification of nonlinear processes under missing observations, The Canadian Journal of Chemical Engineering 86 (6) (2008) 1081-1092.

[17] A.J. Isaksson, Identification of ARX-models subject to missing data, IEEE Transaction Automatic Control 38 (5) (1993) 813-819.

[18] J. Sheng, T. Chen, S.L. Shah, Generalized predictive control for nonuniformly sampled systems, Journal of Process Control 12 (8) (2002) 875-885.

[19] D. Li, S.L. Shah, T. Chen, Identification of fast-rate models from multirate data, International Journal of Control 74 (7) (2001) 680-689.

[20] F. Ding, T. Chen, Combined parameter and output estimation of dual-rate systems using an auxiliary model, Automatica 40 (10) (2004) 1739-1748.

[21] F. Ding, T. Chen, Parameter estimation of dual-rate stochastic systems by using an output error method, IEEE Transaction on Automatic Control 50 (9) (2005) 1436-1441.

[22] F. Ding, P.X. Liu, H. Yang, Parameter identification and intersample output estimation for dual-rate systems, IEEE Transaction on Systems Man and Cybernetics-Part A Systems and Humans 38 (4) (2008) 966-975.

[23] J. Ding, F. Ding, P.X. Liu, G. Liu, Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data, IEEE Transactions on Automatic Control, 56 (11) (2011) 2677-2683.

[24] J. Chen, L.X. Lv, R.F. Ding, Parameter estimation for dual-rate sampled data systems with preload nonlinearities, Advances in Intelligent and Soft Computing 125 (2011) 43-50.

[25] J. Chen, L.X. Lv, R.F. Ding, Multi-innovation stochastic gradient algorithms for dual-rate sampled systems with preload nonlinearity, Applied Mathematics Letters, 26 (1) (2013) 124-129.

[26] H. Raghavan, R.B. Gopaluni, et al., Gray-box identification of dynamic models for the bleaching operation in a pulp mill, Journal of Process Control 15 (4) (2005) 451-468.

[27] P. Stoica, N. Sandgren, Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches, Digital Signal Process, 16 (6) (2006) 712-734.

[28] Y. Wang, P. Stoica, J. Li, T.L. Marzetta, Nonparametric spectral analysis with missing data via the EM algorithm, Digital Signal Process, 15 (2) (2005) 191-206.

[29] G.C. Goodwin, K.S. Sin, Adaptive filtering prediction and control, Englewood Cliffs, New Jersey, Prentice-hall, 1984.

[30] F. Ding, X.P. Liu, G. Liu, Auxiliary model based multi-innovation extended stochastic gradient parameter estimation with colored measurement noises, Signal Processing 89 (10) (2009) 1883-1890.