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

On the residual based stochastic gradient algorithm for dual-rate sampled-data systems using the polynomial transform technique

Yuwu Liao, Dongqing Wang, Xiaoming Chen, Feng Ding

Abstract— **This paper uses the polynomial transformation technique to transform an ARX model into a special model that can be identified with dual-rate input-output data, and presents the residual based stochastic gradient algorithm for dual-rate sampled-data systems, and studies convergence properties of the algorithm involved. The analysis indicates that the parameter estimation error consistently converges to zero under some proper conditions. Finally, we test the algorithms proposed in paper by a simulation example and show their effectiveness.**

I. PROBLEM FORMULATION

Consider an ARX model (auto-regression model with exogenous input) described by

$$A(z)y(t) = B(z)u(t) + v(t), \tag{1}$$

where $\{u(t)\}$ and $\{y(t)\}$ are the system input and output sequences, $\{v(t)\}$ is a random noise sequence with zero mean and unknown time-varying variance σ^2, z^{-1} is a unit backward shift operator, i.e., z is a unit forward shift operator [$z^{-1}u(t) = u(t-1)$, $z^{-1}y(t) = y(t-1)$, $z^{-1}y(qt) = y(qt-1)$, and $z^{-q}y(qt) = y((t-1)q)$], and $A(z)$ and $B(z)$ are polynomials in z^{-1} with

$$A(z) = 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_nz^{-n},$$
$$B(z) = b_1z^{-1} + b_2z^{-2} + \cdots + b_nz^{-n}.$$

Define the information vector $\varphi(t)$ and the parameter vector θ as

$$\varphi(t) = [-y(t-1), \cdots, -y(t-n), u(t-1), \cdots, u(t-n)]^{\mathrm{T}},$$
$$\theta = [a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n]^{\mathrm{T}},$$

where the superscript T denotes the matrix transpose. Equation (1) may be written as

$$y(t) = \varphi^{\mathrm{T}}(t)\theta + v(t), \tag{2}$$

Without loss of generality, assume that $u(t) = 0$, $y(t) = 0$ and $v(t) = 0$ for $t \leqslant 0$, and the system order n is known.

For dual-rate sampled-data systems [5]–[7], all the inputs $\{u(t)\colon\ t = 0, 1, 2, \cdots\}$ are available, but only scarce outputs $\{y(qt)\colon\ t = 0, 1, 2, \cdots\}$ are available ($q \geqslant 2$ is a positive integer). In such systems, the input-output data available are $\{u(t), y(qt)\colon\ t = 0, 1, 2, \cdots\}$. Thus, the intersample outputs (or missing outputs), $y(qt+i)$, $i = 1, 2, \cdots, q-1$, are not available. In a word, the model in (1) is not appropriate for dual-rate system identification. Therefore, (1) needs to be transformed into a form that can be used directly on the dual-rate data. A polynomial transformation technique [3], [4] can be adopted to do this. The details are as follows.

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

$$A(z) = (1 - z_1z^{-1})(1 - z_2z^{-1}) \cdots (1 - z_nz^{-1}).$$

Define a polynomial,

$$\begin{aligned}\phi_q(z) &= \prod_{i=1}^{n}(1 + z_iz^{-1} + z_i^2z^{-2} + \cdots + z_i^{q-1}z^{-q+1}) \\ &=: 1 + d_1z^{-1} + d_2z^{-2} + \cdots + d_mz^{-m}.\end{aligned}$$

Multiplying both sides of (1) by $\phi_q(z)$ and using the formula:

$$1 - x^q = (1 - x)(1 + x + x^2 + \cdots + x^{q-1}).$$

transform the model (1) into the desired form:

$$\alpha(z)y(t) = \beta(z)u(t) + \phi_q(z)v(t) \tag{3}$$

with

$$\begin{aligned}\alpha(z) &= \phi_q(z)A(z) \\ &=: 1 + \alpha_1z^{-q} + \alpha_2z^{-2q} + \cdots + \alpha_nz^{-qn}, \end{aligned} \tag{4}$$
$$\begin{aligned}\beta(z) &= \phi_q(z)B(z) \\ &=: \beta_1z^{-1} + \beta_2z^{-2} + \cdots + \beta_{qn}z^{-qn}. \end{aligned} \tag{5}$$

The objective of this paper is to present an algorithm to estimate the parameters of the model in (3) using available dual-rate input-output data $\{u(t), y(qt)\colon t = 0, 1, 2, \cdots\}$ ($q \geqslant 2$ being an integer), and to study the convergence of the algorithm involved, and to give an illustrative example to show the effectiveness of the algorithms proposed in this paper. Finally, we offer some concluding remarks in Conclusions.

II. THE BASIC ALGORITHMS

To proceed further, let us introduce some notation. The symbol I stands for an identity matrix of appropriate dimensions; the norm of a column vector x is defined as $\|x\|^2 = x^{\mathrm{T}}x$; $\lambda_{\max}[X]$ and $\lambda_{\min}[X]$ represent the maximum and minimum eigenvalues of X, respectively; $\mathbf{1}_n$ represents an n-dimensional vector whose elements are all 1; $f(t) = o(g(t))$ represents $f(t)/g(t) \to 0$ as $t \to \infty$; for $g(t) \geqslant 0$, we write $f(t) = O(g(t))$ if there exists a positive constant δ_1 such that $|f(t)| \leqslant \delta_1 g(t)$.

This work was supported by the Shandong Province Colleges and Universities Outstanding Young Teachers in Domestic Visiting Scholars Project.

Corresponding authors: X.M. Chen and F. Ding are with the Control Science and Engineering Research Center, Jiangnan University, Wuxi, P. R. China 214122. cxmxiaojie@yahoo.com.cn; fding@jiangnan.edu.cn.

Y.W. Liao is with the Department of Physics and Electronics Information Technology, Xiangfan University, Xiangfan, P.R. China 441053. 168.lyw@163.com.

D.Q. Wang is with the College of Automation Engineering, Qingdao University, Qingdao, P.R. China 266071. dqwang64@163.com.

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

Substituting the polynomials $\alpha(z)$ in z^{-q} in (4) and $\beta(z)$ in z^{-1} in (5) into (3) leads to the following regression equation:

$$y(t)=\phi_0^{\mathrm{T}}(t)\vartheta+v(t), \tag{6}$$

where the parameter vector ϑ and information vector $\phi_0(t)$ are defined by

$$\begin{aligned}\vartheta &:= [\alpha_1,\alpha_2,\cdots,\alpha_n,\beta_1,\beta_2,\cdots,\beta_{qn},d_1,d_2,d_m]^{\mathrm{T}}\in\mathbb{R}^{n_0},\\ \phi_0(t) &:= [-y(t-q),-y(t-2q),\cdots,-y(t-qn),\\ &\quad u(t-1),u(t-2),\cdots,u(t-qn),\\ &\quad v(t-1),v(t-2),\cdots,v(t-m)]^{\mathrm{T}}\in\mathbb{R}^{n_0}. \end{aligned}\tag{7}$$

Replacing t with qt in equation (6) gives

$$y(qt)=\phi_0^{\mathrm{T}}(qt)\vartheta+v(qt), \tag{8}$$

$$\begin{aligned}\phi_0(qt) &= [-y(qt-q),-y(qt-2q),\cdots,-y(qt-qn),\\ &\quad u(qt-1),u(qt-2),\cdots,u(qt-qn),\\ &\quad v(qt-1),v(qt-2),\cdots,v(qt-m)]^{\mathrm{T}}.\end{aligned}\tag{9}$$

Let $\hat{\vartheta}(qt)$ be the estimate of ϑ at time qt. Since $v(t)$ is assumed to be a zero-mean random white noise sequence, minimizing the cost function [2],

$$J(\vartheta)=\|y(qt)-\phi_0^{\mathrm{T}}(qt)\vartheta\|^2. \tag{10}$$

The unknowns $v(qt-i)$, $i=1,2,\cdots,qn$, in $\phi_0(qt)$ are replaced with their estimates $\hat{v}(qt-i)$, $\phi_0(qt)$ with $\phi(qt)$, then applying the negative gradient search principle to (10), one can get the following recursive stochastic gradient algorithm of estimating ϑ in (6) based on the noise estimation (The DR-RESG algorithm for short):

$$\hat{\vartheta}(qt)=\hat{\vartheta}(qt-q)+\frac{\phi(qt)}{r(qt)}\times [y(qt)-\phi^{\mathrm{T}}(qt)\hat{\vartheta}(qt-q)], \tag{11}$$

$$\hat{\vartheta}(qt+i)=\hat{\vartheta}(qt),\ i=0,1,\cdots,q-1, \tag{12}$$

$$r(qt)=r(qt-q)+\|\phi(qt)\|^2,\ r(0)=1, \tag{13}$$

$$\begin{aligned}\phi(qt) &= [-y(qt-q),-y(qt-2q),\cdots,-y(qt-qn),\\ &\quad u(qt-1),u(qt-2),\cdots,u(qt-qn),\\ &\quad \hat{v}(qt-1),\hat{v}(qt-2),\cdots,\hat{v}(qt-m)]^{\mathrm{T}},\end{aligned}\tag{14}$$

$$\hat{v}(qt-i)=\hat{y}(qt-i)-\phi^{\mathrm{T}}(qt-i)\hat{\vartheta}(qt), \tag{15}$$

However, when i is not an integer multiple of q, $y(qt-i)$ and $\phi^{\mathrm{T}}(qt-i)$ in (15) involve missing outputs, it is not feasible to compute the residuals by (15). The solution is to the missing outputs $y(qt+i)$ are replaced with their estimates $\hat{y}(qt+i)$. The following is to give a way to compute $\hat{y}(qt+i)$.

Using the obtained $\hat{\vartheta}(qt)$ to form the polynomials,

$$\begin{aligned}\hat{\alpha}(qt,z) &= 1+\hat{\alpha}_1(qt)z^{-q}+\hat{\alpha}_2(qt)z^{-2q}+\cdots+\hat{\alpha}_n(qt)z^{-qn},\\ \hat{\beta}(qt,z) &= \hat{\beta}_1(qt)z^{-1}+\hat{\beta}_2(qt)z^{-2}+\cdots+\hat{\beta}_{qn}(qt)z^{-qn}.\end{aligned}$$

Dividing both sides of (5) by both sides of (4) gives

$$\frac{B(z)}{A(z)}=\frac{\beta(z)}{\alpha(z)}.$$

Assume that the estimates of $A(z)$ and $B(z)$ at time qt are

$$\begin{aligned}\hat{A}(qt,z) &= 1+\hat{a}_1(qt)z^{-1}+\hat{a}_2(qt)z^{-2}+\cdots+\hat{a}_n(qt)z^{-n},\\ \hat{B}(qt,z) &= \hat{b}_1(qt)z^{-1}+\hat{b}_2(qt)z^{-2}+\cdots+\hat{b}_{qn}(qt)z^{-n}.\end{aligned}$$

According to the model equivalence principle in [8], [9], let

$$\frac{\hat{B}(qt,z)}{\hat{A}(qt,z)}=\frac{\hat{\beta}(qt,z)}{\hat{\alpha}(qt,z)},$$

or

$$\hat{\alpha}(qt,z)\hat{B}(qt,z)=\hat{\beta}(qt,z)\hat{A}(qt,z).$$

One can compute $\hat{A}(qt,z)$ and $\hat{B}(qt,z)$ by comparing the coefficients of z^{-i} in both sides. Then the missing outputs are computed by

$$\hat{y}(qt+i)=\begin{cases} y(qt), & i=0,\\ \frac{\hat{B}(qt,z)}{\hat{A}(qt,z)}u(qt+i), & i=1,2,\cdots,q-1.\end{cases}$$

Or

$$\hat{y}(qt+i)=\begin{cases} y(qt), & i=0,\\ \hat{\varphi}^{\mathrm{T}}(qt+i)\hat{\theta}(qt), & i=1,2,\cdots,q-1,\end{cases}\tag{16}$$

where

$$\begin{aligned}\hat{\varphi}(qt+i) &:= [-\hat{y}(qt+i-1),-\hat{y}(qt+i-2),\cdots,\\ &\quad -\hat{y}(qt+i-n),u(qt+i-1),\\ &\quad u(qt+i-2),\cdots,u(qt+i-n)]^{\mathrm{T}},\end{aligned}\tag{17}$$

$$\hat{\theta}(qt):=[\hat{a}_1(qt),\cdots,\hat{a}_n(qt),\hat{b}_1(qt),\cdots,\hat{b}_n(qt)]^{\mathrm{T}}. \tag{18}$$

To initialize the DR-RESG algorithm in (11)-(18), we still take $\hat{\vartheta}(0)=10^{-6}\mathbf{1}_n$ with $\mathbf{1}_{n_0}$ being an n_0-dimensional vector whose elements are all 1.

III. Convergence of the DR-RESG algorithm

We assume that $\{v(t),\mathscr{F}_t\}$ is a martingale difference sequence defined on a probability space $\{\Omega,\mathscr{F},P\}$, where $\{\mathscr{F}_t\}$ is the σ algebra sequence generated by $\{v(t)\}$, i.e., $\mathscr{F}_t=\sigma(v(t),v(t-1),v(t-2),\cdots,)$ or $\mathscr{F}_t=\sigma(y(t),y(t-1),y(t-2),\cdots,)$ for the deterministic input sequence $\{u(t)\}$. The noise sequence $\{v(t)\}$ satisfies the following assumptions [1]:

$$\begin{aligned}&(A1)\ \mathrm{E}[v(t)|\mathscr{F}_{t-1}]=0,\ \text{a.s.},\\ &(A2)\ \mathrm{E}[v^2(t)|\mathscr{F}_{t-1}]\leqslant\sigma^2<\infty,\ \text{a.s.}\end{aligned}$$

Theorem 1: For the system in (8) and the DR-RESG algorithm in (11)-(18), define

$$R(qt):=\sum_{i=1}^{t}\phi(iq)\phi^{\mathrm{T}}(iq),$$

and assume that (A1) and (A2) hold, $\phi_q(z)$ is strictly positive real, and $r(qt)=O(\lambda_{\min}[R(qt)])$. Then the parameter estimation vector consistently converges to the true parameter vector ϑ, i.e., $\hat{\vartheta}(qt)\to\vartheta$ as $t\to\infty$.

The proof is omitted but available from the authors.

The DR-RESG algorithm has low computational effort, but its convergence is relatively slow. In order to improve the tracking performance of the DR-RESG algorithm,

TABLE I

THE DR-RESG ESTIMATES AND ERRORS OF ϑ ($\sigma^2 = 0.10^2$, $\delta_{ns} = 14.69\%$)

Algorithms	t	α_1	α_2	β_1	β_2	β_3	β_4	δ (%)
DR-RESG	100	-0.49904	0.19824	0.37408	0.44907	0.55901	0.25481	47.99108
(DR-REFG,	200	-0.52961	0.22990	0.37345	0.46261	0.57443	0.25936	45.40041
	300	-0.55341	0.25034	0.37748	0.47276	0.58302	0.26687	43.58943
$\lambda = 1$)	500	-0.56938	0.26039	0.37903	0.47941	0.59099	0.27037	42.45893
	800	-0.58395	0.27254	0.38109	0.48687	0.59825	0.27168	41.29148
	1000	-0.59230	0.28303	0.38229	0.49185	0.60226	0.27261	40.50002
	1500	-0.60586	0.29765	0.38311	0.49972	0.60803	0.27584	39.31489
	2000	-0.61450	0.30761	0.38342	0.50569	0.61179	0.27836	38.51016
	2500	-0.62196	0.31531	0.38427	0.50997	0.61530	0.27977	37.86735
	3000	-0.62646	0.32139	0.38419	0.51332	0.61732	0.28049	37.41425
DR-REFG	100	-0.52383	0.22263	0.37416	0.46502	0.56999	0.26343	45.78888
$\lambda = 0.99$	200	-0.57301	0.27605	0.37363	0.48849	0.59512	0.26939	41.51881
	300	-0.62066	0.32031	0.38283	0.51154	0.61315	0.28363	37.75897
	500	-0.67123	0.34931	0.38793	0.53840	0.64005	0.29087	34.11091
	800	-0.72882	0.41140	0.39662	0.58468	0.67443	0.28648	28.51938
	1000	-0.75980	0.46710	0.40122	0.61830	0.69551	0.28299	24.61997
	1500	-0.82117	0.52608	0.40562	0.68186	0.72838	0.28507	18.58141
	2000	-0.85632	0.56332	0.40429	0.73316	0.75228	0.28500	14.40285
	2500	-0.88665	0.57999	0.40515	0.77219	0.77212	0.27949	11.33927
	3000	-0.89417	0.60153	0.40422	0.80785	0.78281	0.26981	8.92581
DR-REFG	100	-0.56924	0.24993	0.42555	0.54992	0.64804	0.18925	39.66257
$\lambda = 0.98$	200	-0.64603	0.34228	0.41958	0.57771	0.67719	0.20702	33.02145
	300	-0.71239	0.39426	0.42287	0.61048	0.69488	0.23046	28.06217
	500	-0.78220	0.45988	0.42804	0.65139	0.72597	0.24729	22.36451
	800	-0.84361	0.53107	0.42723	0.72047	0.75094	0.24032	15.80564
	1000	-0.88165	0.57533	0.42270	0.75911	0.76597	0.24463	12.06676
	1500	-0.92000	0.61053	0.42148	0.82001	0.78618	0.25060	7.57510
	2000	-0.93767	0.62726	0.41222	0.86042	0.79484	0.24559	4.84145
	2500	-0.94339	0.63274	0.41012	0.89314	0.79678	0.24026	2.92962
	3000	-0.94600	0.62855	0.40698	0.90659	0.79894	0.23806	2.20742
True values		-0.96000	0.64000	0.40000	0.94000	0.80000	0.24000	

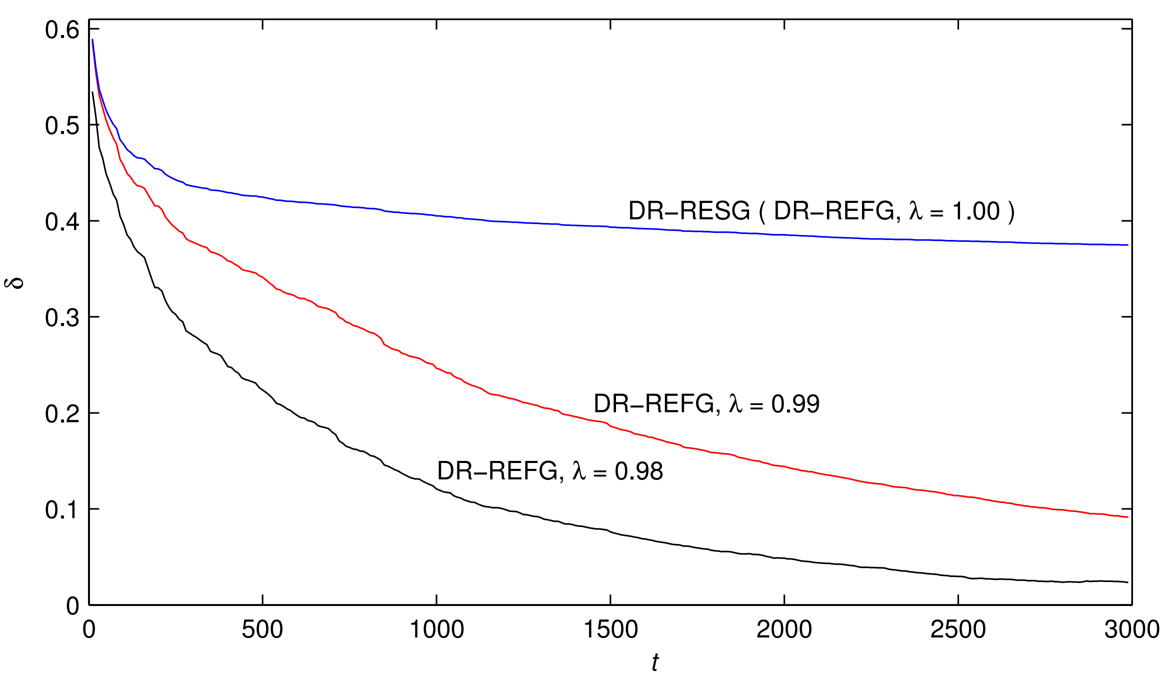

Fig. 1. The estimation errors δ v.s. t ($\sigma^2 = 0.10^2$, $\delta_{ns} = 14.69\%$)

TABLE II

THE ESTIMATES OF a_i AND b_i ($\sigma^2 = 0.10^2$, $\delta_{\rm ns} = 14.69\%$)

Algorithms	t	a_1	a_2	b_1	b_2	δ_s (%)
DR-RESG	100	-1.07982	0.34227	0.19917	0.12402	39.97844
(DR-REFG,	200	-1.10629	0.37042	0.18747	0.13500	38.09528
	300	-1.12457	0.38753	0.18228	0.13727	36.91137
$\lambda = 1$)	500	-1.13334	0.39556	0.17788	0.14162	36.34654
	800	-1.14244	0.40416	0.17517	0.14625	35.72194
	1000	-1.15019	0.41198	0.17390	0.14867	35.17724
	1500	-1.16585	0.42727	0.17184	0.15192	34.10955
	2000	-1.17855	0.43966	0.17113	0.15409	33.23784
	2500	-1.18541	0.44624	0.17017	0.15604	32.77407
	3000	-1.19248	0.45325	0.17000	0.15760	32.27972
DR-REFG	100	-1.11897	0.38126	0.19350	0.12987	37.23482
$\lambda = 0.99$	200	-1.16110	0.42514	0.17807	0.14870	34.27505
	300	-1.19729	0.45856	0.17121	0.15381	31.96333
	500	-1.22912	0.48555	0.16517	0.17059	29.89469
	800	-1.28371	0.53527	0.17325	0.19567	25.95538
	1000	-1.32904	0.57961	0.18550	0.20981	22.59519
	1500	-1.41685	0.65737	0.21592	0.23116	16.37515
	2000	-1.48131	0.71466	0.24759	0.24897	11.69733
	2500	-1.51419	0.73922	0.27199	0.26441	9.12259
	3000	-1.54258	0.76569	0.30165	0.27482	6.54527
DR-REFG	100	-1.04000	0.30437	0.22257	0.21809	41.61397
$\lambda = 0.98$	200	-1.12364	0.38507	0.19988	0.22656	35.89552
	300	-1.21106	0.46240	0.19957	0.22527	30.02458
	500	-1.28348	0.52698	0.20165	0.23519	25.15288
	800	-1.39087	0.62372	0.24352	0.25337	17.14992
	1000	-1.45793	0.68472	0.26776	0.26079	12.33582
	1500	-1.53330	0.75086	0.31074	0.27066	6.74400
	2000	-1.57013	0.78133	0.34292	0.28297	3.72561
	2500	-1.58628	0.79382	0.36979	0.28821	1.92467
	3000	-1.58740	0.79184	0.37897	0.29453	1.42210
True values		-1.60000	0.80000	0.40000	0.30000	

we introduce a forgetting factor λ in the DR-RESG algorithm in (11)-(15) to get

$$\hat{\vartheta}(qt) = \hat{\vartheta}(qt-q) + \frac{\phi(qt)}{r(qt)} \times [y(qt) - \phi^{\rm T}(qt)\hat{\vartheta}(qt-q)], \quad (19)$$

$$\hat{\vartheta}(qt+i) = \hat{\vartheta}(qt),\ i = 0, 1, \cdots, q-1, \quad (20)$$

$$r(qt) = \lambda r(qt-q) + \|\phi(qt)\|^2,\ 0 \leqslant \lambda \leqslant 1,\ r(0) = 1, \quad (21)$$

$$\phi(qt) = [-y(qt-q), -y(qt-2q), \cdots, -y(qt-qn), u(qt-1), u(qt-2), \cdots, u(qt-qn), \hat{v}(qt-1), \hat{v}(qt-2), \cdots, \hat{v}(qt-m)]^{\rm T}, \quad (22)$$

$$\hat{v}(qt-i) = \hat{y}(qt-i) - \phi^{\rm T}(qt-i)\hat{\vartheta}(qt), \quad (23)$$

Equation (19)-(23) form the DR-RESG algorithm with a forgetting factor (refer to as the DR-REFG algorithm). When $\lambda = 1$, the DR-REFG algorithm reduces to the DR-RESG algorithm; when $\lambda = 0$, the DR-REFG algorithm is the dual-rate projection algorithm.

IV. EXAMPLE

Example consider a second-order discrete-time system with

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} = 1 - 1.60z^{-1} + 0.80z^{-2},$$
$$B(z) = b_1 z^{-1} + b_2 z^{-2} = 0.40z^{-1} + 0.30z^{-2}.$$

Here, we take $q = 2$. The corresponding dual-rate model can be expressed as

$$\alpha(z)y(t) = \beta(z)u(t) + \phi_q(z)v(t),$$
$$\alpha(z) = \phi_q(z)a(z),\ \beta(z) = \phi_q(z)B(z),$$
$$\phi_q(z) = 1 - d_1 z^{-1} + d_2 z^{-2} = 1 + 1.60z^{-1} + 0.80z^{-2}.$$

Here $\{u(t)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance, and $\{v(t)\}$ as a white noise sequence with zero mean and variance σ^2. Applying the DR-REFG algorithm with the forgetting factors to estimate the parameters (α_i, β_i) of this system, and and using the approach given in [8] to compute the estimates of a_i and b_i, the parameter estimates and their estimation errors $\delta := \|\hat{\vartheta}(t) - \vartheta\|/\|\vartheta\|$ or $\delta_s := \|\hat{\theta}(t) - \theta\|/\|\theta\|$ are shown in Table I and Table II, and the estimation errors δ versus t are shown in Figure 1 with $\lambda = 1.00$, $\lambda = 0.99$ and $\lambda = 0.98$, where the noise variance $\sigma^2 = 0.10^2$ and corresponding noise-to-signal ratio $\delta_{\rm ns} = 14.69\%$.

From Table I and Figure 1, we can see that the forgetting factor λ is increased, the rate of change of the parameter estimates becomes more stationary, but the estimation error gets larger; the error δ is becoming smaller (in general) as t increases.

V. CONCLUSIONS

The performance of DR-RESG algorithms is studied for dual-rate sampled-data systems based on ARX models; the analysis indicates that the algorithms proposed can achieve good performance properties and require less computational efforts than the exiting algorithm. Although the analysis method used in the paper is done for dual-rate models with a colored noise, the methods developed can be easily extended to dual-rate stochastic systems with additive white noises.

REFERENCES

[1] G.C. Goodwin, K.S. Sin, Adaptive Filtering Prediction and Control, Englewood Cliffs, NJ: Prentice-hall, 1984.

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

[3] F. Ding, H. Chen, Y. Xiao, "Consistency of stochastic gradient parameter and missing output estimation by using a polynomial transformation," *Dynamics of Continuous Discrete and Impulsive Systems-Series A-Mathematical Analysis*, vol. 13 (Suppl. Part 3), pp. 1308-1313.

[4] F. Ding, P.X. Liu, H.Z. Yang, "Parameter Identification and Intersample Output Estimation for Dual-Rate Systems," *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, vol. 38, no. 4, pp. 966-975, 2008.

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

[6] F. Ding, T. Chen, "Parameter estimation of dual-rate stochastic systems by using an output error method," *IEEE Transactions on Automatic Control*, vol. 50, no. 9, pp. 1436-1441, 2005.

[7] F. Ding, 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, 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.

[9] F. Ding, J.B. Yang, "Two model simplification methods of large-scale systems," *Systems Engineering*, vol. 16, no. 5, pp. 60-66, 1998.