Hierarchical Least Squares Identification for Linear SISO Systems With Dual-Rate Sampled-Data

Jie Ding, Feng Ding, Xiaoping Peter Liu, and Guangjun Liu

***Abstract*—This technical note studies identification problems for dual-rate sampled-data linear systems with noises. A hierarchical least squares (HLS) identification algorithm is presented to estimate the parameters of the dual-rate ARMAX models. The basic idea is to decompose the identification model of a dual-rate system into several sub-identification models with smaller dimensions and fewer parameters. The proposed algorithm is more computationally efficient than the recursive least squares (RLS) algorithm since the RLS algorithm requires computing the covariance matrix of large sizes, while the HLS algorithm deals with the covariance matrix of small sizes. Compared with our previous work, a detailed study of the HLS algorithm is conducted in this technical note. The performance analysis and simulation results confirm that the estimation accuracy of the proposed algorithm are close to that of the RLS algorithm, but the proposed algorithm retains much less computational burden.**

***Index Terms*—Convergence, dual-rate systems, hierarchical identification, least squares, parameter estimation.**

I. Introduction

In the process industries, all the variables are impossible to be sampled at the same rate due to the sensor limits. For instance, chemical process variables such as temperature, pressure and flow rate can be readily sampled at a high rate by sensors, while quality variables as the composition, density and molecular weight distribution can be acquired only at a low frequency [1], [2]. Systems operating at different input and output sampling rates are called multirate systems [3]–[6] which have wide applications in control, communication, signal processing, etc.

In the literature of system identification, lots of attention has been paid to dual-rate/multirate systems [7]–[10]. This technical note considers identification problems of a class of dual-rate systems, where the sampling period of a process output is an integer multiple of the holding period of a control input. That is, all the inputs $\{u(k), k = 0, 1, 2, \cdots\}$ are available at each sampling time k, but only scarce outputs $\{y(kq), k = 0, 1, 2, \cdots\}$ ($q > 1$ is an integer) are measurable, thus, the available input-output data are $\{u(k), y(kq) : k = 0, 1, 2, \cdots\}$.

The identification methods for the dual-rate sampled-data systems can be largely divided into two categories: the lifting technique based state-space estimation methods [11]–[14] and the polynomial transformation technique based estimation methods [15]–[17]. By using the lifting technique, the fast input and slow output data are collected and a multivariable state-space model can be identified; by using the polynomial transformation technique, a dual-rate model can be estimated directly by utilizing the fast input and slow output data.

Since the output $y(kq)$ is sampled at a slower rate than the input $u(k)$, the intersample outputs $\{y(kq + j), j = 1, 2, \cdots, q - 1\}$ are missing. In this case, Shumway and Stoffer considered linear state-space models and used an expectation maximization (EM) algorithm to handle the missing measurements [18]. These results have been extended to handle the nonlinear state-space models in [19], [20]. Isaksson studied identification problems of ARX models with missing data based on the Kalman filtering (fixed-interval smoothing) technique and maximum likelihood (ML) methods [21]. The EM and ML approaches can off-line tackle the dual-rate or multirate problems but cannot work in the on-line setting as treated in this technical note.

On the basis of the work in [22] which handled a dual-rate ARX model from dual-rate data, this technical note deals with identification problems of dual-rate ARMAX systems with moving average noise. The main contributions of this technical note are to develop a new hierarchical least squares algorithm for such dual-rate systems and to discuss how to choose the number of fictitious subsystems and the advantages of the proposed algorithm compared with the RLS algorithm.

The rest of this technical note is organized as follows. Section II derives a dual-rate system identification model by using a polynomial transformation technique. Based on this model, Section III develops a hierarchical least squares identification algorithm. Section IV proves the convergence properties of the parameter estimation given by the proposed algorithm. Section V gives an illustrative example to show the effectiveness of the proposed algorithm in the technical note. Finally, we offer some concluding remarks in Section VI.

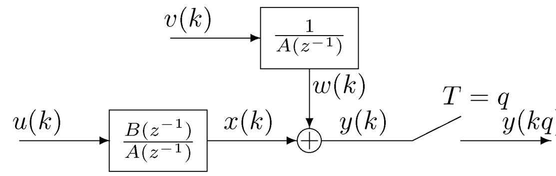

Fig. 1. Dual-rate sampled-data system with noises.

II. Problem Formulation

Consider a dual-rate sampled-data system in Fig. 1, where $\{u(k)\}$ and $\{y(k)\}$ are the system input and output sequences with the same sampling period h, $\{v(k)\}$ is a white noise sequence with zero mean and variance σ^2, $y(kq)$ is the slow rate sampled output processed by a discrete-time sampler with sampling period $T = q$, $A(z^{-1})$ and $B(z^{-1})$ are the polynomials in z^{-1} (z^{-1} is the unit backward shift operator, i.e., $z^{-1}u(k) = u(k-1)$)

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

the parameters a_i and b_i are unknown, the orders n_a and n_b are assumed to be known. From Fig. 1, we have the single-rate (fast rate) model [23]

$$A(z^{-1})y(k) = B(z^{-1})u(k) + v(k). \tag{1}$$

Without loss of generality, assume that $u(k) = 0$, $y(k) = 0$ and $v(k) = 0$ for $k \leqslant 0$.

Manuscript received May 12, 2010; revised December 18, 2010, May 04, 2011, and May 05, 2011; accepted May 19, 2011. Date of publication May 31, 2011; date of current version November 02, 2011. Recommended by Associate Editor P. Shi.

J. Ding and F. Ding are with the Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China (e-mail: djvsqxb@163.com; fding@jiangnan.edu.cn).

X. P. Liu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: xpliu@sce.carleton.ca).

G. Liu is with the Department of Aerospace Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: gjliu@ryerson.ca).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2011.2158137

Define the parameter vector $\boldsymbol{\theta}$ and the regressor $\boldsymbol{\varphi}(k)$ as

$$\begin{aligned}\boldsymbol{\theta} &= [a_1, a_2, \cdots, a_{n_a}, b_0, b_1, b_2, \cdots, b_{n_b}]^{\mathrm{T}} \in \mathbb{R}^{n_0},\\ \boldsymbol{\varphi}(k) &= [-y(k-1), -y(k-2), \cdots, -y(k-n_a),\\ &\qquad u(k), u(k-1), u(k-2), \cdots, u(k-n_b)]^{\mathrm{T}} \in \mathbb{R}^{n_0},\\ n_0 &= n_a + n_b + 1.\end{aligned}$$

Then (1) can be rewritten in a regressive form

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

Notice that the available output in Fig. 1 is $y(kq)$, that is, the regressor $\boldsymbol{\varphi}(k)$ contains the missing outputs $y(kq-l)$, when l is not a multiple of q.

The polynomial transformation technique is applied to (1) to obtain a model that can directly use the dual-rate sampled data $\{u(k), y(kq)\}$ as in [16], [24]. Let the roots of $A(z^{-1})$ be z_l $(l = 1, 2, \cdots, n_a)$ to get

$$A(z^{-1}) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) \cdots \left(1 - z_{n_a} z^{-1}\right).$$

Introducing the polynomials

$$\begin{aligned}D(z^{-1}) &= \prod_{l=1}^{n_a} \left(1 + z_l z^{-1} + z_l^2 z^{-2} + \cdots + z_l^{q-1} z^{-q+1}\right)\\ &= 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots + d_{n_d} z^{-n_d},\\ \alpha(z^{-1}) &= D(z^{-1})A(z^{-1})\\ &= 1 + \alpha_1 z^{-q} + \alpha_2 z^{-2q} + \cdots + \alpha_{n_a} z^{-n_a q} \qquad (3)\\ \beta(z^{-1}) &= D(z^{-1})B(z^{-1})\\ &= \beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \cdots + \beta_m z^{-m},\\ n_d &= (q-1)n_a, \quad m = n_d + n_b. \qquad (4)\end{aligned}$$

Multiplying both sides of (1) by $D(z^{-1})$ gives

$$\alpha(z^{-1})y(k) = \beta(z^{-1})u(k) + D(z^{-1})v(k). \tag{5}$$

Notice that $\alpha(z^{-1})$ and $\beta(z^{-1})$ are polynomials in z^{-q} and z^{-1}, respectively. Thus the above model can be identified directly by using the dual-rate sampled data $\{u(k), y(kq)\}$.

Define the parameter vector $\boldsymbol{\vartheta}$ and the regressor $\boldsymbol{\phi}(k)$ as

$$\begin{aligned}\boldsymbol{\vartheta} &= [\alpha_1, \alpha_2, \cdots, \alpha_{n_a}, \beta_0, \beta_1, \beta_2, \cdots, \beta_m,\\ &\qquad d_1, d_2, \cdots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^n, \quad n = n_a + m + 1 + n_d,\\ \boldsymbol{\phi}(k) &= [-y(k-q), -y(k-2q), \cdots, -y(k-n_a q),\\ &\qquad u(k), u(k-1), u(k-2), \cdots, u(k-m),\\ &\qquad v(k-1), v(k-2), \cdots, v(k-n_d)]^{\mathrm{T}} \in \mathbb{R}^n.\end{aligned}$$

Equation (5) can be rewritten in a regressive form $y(k) = \boldsymbol{\phi}^{\mathrm{T}}(k)\boldsymbol{\vartheta} + v(k)$. Replacing k with kq gives

$$y(kq) = \boldsymbol{\phi}^{\mathrm{T}}(kq)\boldsymbol{\vartheta} + v(kq). \tag{6}$$

From (6), we can see that the regressor $\boldsymbol{\phi}(kq)$ contains not only the available dual-rate data $\{u(k), y(kq) : k = 0, 1, 2, \cdots\}$, but also the unmeasurable noise terms $v(kq-l)$, thus the RLS algorithm [25], [26] is applied to generate the estimate $\hat{\boldsymbol{\vartheta}}(kq)$ of $\boldsymbol{\vartheta}$ as follows:

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}(kq) &= \hat{\boldsymbol{\vartheta}}(kq-q) + \boldsymbol{P}(kq)\hat{\boldsymbol{\phi}}(kq)\\ &\quad \times \left[y(kq) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}(kq-q)\right]\end{aligned} \tag{7}$$

$$\hat{\boldsymbol{\vartheta}}(kq+j) = \hat{\boldsymbol{\vartheta}}(kq), \quad j = 1, 2, \cdots, q-1 \tag{8}$$

$$\begin{aligned}\boldsymbol{P}(kq) &= \boldsymbol{P}(kq-q)\\ &\quad - \frac{\boldsymbol{P}(kq-q)\hat{\boldsymbol{\phi}}(kq)\hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\boldsymbol{P}(kq-q)}{1 + \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\boldsymbol{P}(kq-q)\hat{\boldsymbol{\phi}}(kq)},\end{aligned} \tag{9}$$

$$\begin{aligned}\hat{\boldsymbol{\phi}}(kq) &= [-y(kq-q), -y(kq-2q), \cdots, -y(kq-n_a q),\\ &\qquad u(kq), u(kq-1), u(kq-2), \cdots, u(kq-m),\\ &\qquad \hat{v}(kq-1), \hat{v}(kq-2), \cdots, \hat{v}(kq-n_d)]^{\mathrm{T}},\end{aligned} \tag{10}$$

$$\hat{v}(kq-l) = \hat{y}(kq-l) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq-l)\hat{\boldsymbol{\vartheta}}(kq-q), \tag{11}$$

$$\begin{aligned}\hat{\boldsymbol{\phi}}(kq-l) &= \left[\hat{\boldsymbol{\psi}}^{\mathrm{T}}(kq-l), \hat{v}(kq-l-1), \hat{v}(kq-l-2),\right.\\ &\qquad \left.\cdots, \hat{v}(kq-l-n_d)\right]^{\mathrm{T}},\end{aligned} \tag{12}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}(kq-l) &= [-\hat{y}(kq-l-q), -\hat{y}(kq-l-2q), \cdots,\\ &\qquad -\hat{y}(kq-l-n_a q), u(kq-l), u(kq-l-1),\\ &\qquad u(kq-l-2), \cdots, u(kq-l-m)]^{\mathrm{T}}\end{aligned} \tag{13}$$

where $\hat{y}(kq+j)$ $(j = 1, 2, \cdots, q-1)$ can be computed by linear interpolation of $y(kq)$ and $y(kq+q)$. The initial values are taken to be $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_n/p_0$ and $\boldsymbol{P}(0) = p_0\boldsymbol{I}$ with p_0 normally a large positive number (e.g., $p_0 = 10^6$), $\mathbf{1}_n$ an n-dimensional column vector whose elements are 1 and $\boldsymbol{I}$ an identity matrix with proper size.

Because the RLS algorithm in (7)–(13) requires computing the covariance matrix $\boldsymbol{P}(kq)$ of large sizes, especially for large $n = \dim \boldsymbol{\vartheta}$, developing computationally efficient hierarchical identification algorithms is the goal of this technical note.

III. Hierarchical Least Squares Algorithm

The hierarchical identification principle is employed to ease the computational burden and contains three steps [13], [27].

The First Step: Decomposition: The system in (6) is decomposed into N subsystems, consequently, the regressor $\boldsymbol{\phi}(kq)$ in (6) is decomposed into N sub-regressors $\boldsymbol{\phi}_i(kq)$ with dimensions n_i and the parameter vector $\boldsymbol{\vartheta}$ into N sub-parameter vectors $\boldsymbol{\vartheta}_i$ with dimensions n_i, respectively, i.e.

$$\begin{aligned}\boldsymbol{\phi}(kq) &= \left[\boldsymbol{\phi}_1^{\mathrm{T}}(kq), \boldsymbol{\phi}_2^{\mathrm{T}}(kq), \cdots, \boldsymbol{\phi}_N^{\mathrm{T}}(kq)\right]^{\mathrm{T}} \in \mathbb{R}^n,\\ \boldsymbol{\vartheta} &= \left[\boldsymbol{\vartheta}_1^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}}, \cdots, \boldsymbol{\vartheta}_N^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^n\end{aligned}$$

where $\boldsymbol{\phi}_i(kq) \in \mathbb{R}^{n_i}$, $\boldsymbol{\vartheta}_i \in \mathbb{R}^{n_i}$ $(n_1 + n_2 + \cdots + n_N = n)$. Let

$$y_i(kq) = y(kq) - \sum_{j=1, j\neq i}^{N} \boldsymbol{\phi}_j^{\mathrm{T}}(kq)\boldsymbol{\vartheta}_j, \quad i = 1, 2, \cdots, N. \tag{14}$$

From (6), we can get N subsystems (sub-identification models)

$$y_i(kq) = \boldsymbol{\phi}_i^{\mathrm{T}}(kq)\boldsymbol{\vartheta}_i + v(kq), \quad i = 1, 2, \cdots, N. \tag{15}$$

However, there exist the associate terms $\boldsymbol{\vartheta}_i$ and $\boldsymbol{\vartheta}_j$ between the ith and jth sub-identification models for $i \neq j$. In other words, the ith subsystem in (15) contains the unknown parameter vectors $\boldsymbol{\vartheta}_j$ $(j \neq i)$ of other subsystems.

The Second Step: Subsystem Identification: To minimize the cost function and derive the identification algorithm of each subsystem according to the least squares principle.

Let $\hat{\boldsymbol{\vartheta}}_i(kq)$ be the estimate of the ith parameter vector $\boldsymbol{\vartheta}_i(kq)$. For the sub-identification model in (15), the RLS algorithm for estimating $\boldsymbol{\vartheta}_i$ can be expressed as

$$\hat{\boldsymbol{\vartheta}}_i(kq) = \hat{\boldsymbol{\vartheta}}_i(kq-q) + \boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq) \times \left[y_i(kq) - \hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}_i(kq-q)\right] \tag{16}$$

$$\hat{\boldsymbol{\vartheta}}_i(kq+j) = \hat{\boldsymbol{\vartheta}}_i(kq), \quad j = 0, 1, \cdots, q-1 \tag{17}$$

$$\boldsymbol{P}_i(kq) = \boldsymbol{P}_i(kq-q) - \frac{\boldsymbol{P}_i(kq-q)\hat{\boldsymbol{\phi}}_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq-q)}{1 + \hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq-q)\hat{\boldsymbol{\phi}}_i(kq)}, \tag{18}$$

$$\hat{\boldsymbol{\phi}}(kq) = \left[\hat{\boldsymbol{\phi}}_1^{\mathrm{T}}(kq), \hat{\boldsymbol{\phi}}_2^{\mathrm{T}}(kq), \cdots, \hat{\boldsymbol{\phi}}_N^{\mathrm{T}}(kq)\right]^{\mathrm{T}} \tag{19}$$

where $\boldsymbol{P}_i(kq)$ is the covariance matrix of the ith subsystem. Substituting (14) into (16) gives

$$\hat{\boldsymbol{\vartheta}}_i(kq) = \hat{\boldsymbol{\vartheta}}_i(kq-q) + \boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq) \times \left[y(kq) - \sum_{j=1, j\neq i}^{N} \hat{\boldsymbol{\phi}}_j^{\mathrm{T}}(kq)\boldsymbol{\vartheta}_j - \hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}_i(kq-q)\right]. \tag{20}$$

Here, a difficulty arises in that the expression on the right-hand side of the above equation contains the unknown parameter vector $\boldsymbol{\vartheta}_j$ $(j \neq i)$, so the algorithm in (16)–(20) is impossible to implement. The solution here is the coordination based on the hierarchical identification principle [13], [27].

The Third Step: Coordination: To deal with the associate items among the subsystems.

A coordination way is that the unknown vectors $\boldsymbol{\vartheta}_j$ $(j \neq i)$ of other subsystems which appeared in the ith subsystem in (20) are replaced with their corresponding estimates $\hat{\boldsymbol{\vartheta}}_j(kq-q)$ at the preceding time $kq-q$, so we have

$$\hat{\boldsymbol{\vartheta}}_i(kq) = \hat{\boldsymbol{\vartheta}}_i(kq-q) + \boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq) \times \left[y(kq) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}(kq-q)\right], \quad i = 1, 2, \cdots, N \tag{21}$$

where $\hat{\boldsymbol{\vartheta}}(kq) = [\hat{\boldsymbol{\vartheta}}_1^{\mathrm{T}}(kq), \hat{\boldsymbol{\vartheta}}_2^{\mathrm{T}}(kq), \cdots, \hat{\boldsymbol{\vartheta}}_N^{\mathrm{T}}(kq)]^{\mathrm{T}}$ is the estimate of $\boldsymbol{\vartheta} = [\boldsymbol{\vartheta}_1^{\mathrm{T}}, \boldsymbol{\vartheta}_2^{\mathrm{T}}, \cdots, \boldsymbol{\vartheta}_N^{\mathrm{T}}]^{\mathrm{T}}$. Equations (21), (17), (18) and (19) form the HLS identification algorithm of estimating $\boldsymbol{\vartheta}$ for dual-rate systems (the DR-HLS algorithm for short) and are summarized as follows:

$$\hat{\boldsymbol{\vartheta}}_i(kq) = \hat{\boldsymbol{\vartheta}}_i(kq-q) + \boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq) \times \left[y(kq) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}(kq-q)\right] \tag{22}$$

$$\hat{\boldsymbol{\vartheta}}_i(kq+j) = \hat{\boldsymbol{\vartheta}}_i(kq), \quad j = 1, 2, \cdots, q-1 \tag{22}$$

$$\boldsymbol{P}_i(kq) = \boldsymbol{P}_i(kq-q) - \frac{\boldsymbol{P}_i(kq-q)\hat{\boldsymbol{\phi}}_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq-q)}{1 + \hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq-q)\hat{\boldsymbol{\phi}}_i(kq)}, \tag{24}$$

$$\begin{aligned}\hat{\boldsymbol{\phi}}(kq) &= [-y(kq-q), -y(kq-2q), \cdots, -y(kq-n_a q), \\ &\quad u(kq), u(kq-1), u(kq-2), \cdots, u(kq-m), \\ &\quad \hat{v}(kq-1), \hat{v}(kq-2), \cdots, \hat{v}(kq-n_d)]^{\mathrm{T}} \\ &= \left[\hat{\boldsymbol{\phi}}_1^{\mathrm{T}}(kq), \hat{\boldsymbol{\phi}}_2^{\mathrm{T}}(kq), \cdots, \hat{\boldsymbol{\phi}}_N^{\mathrm{T}}(kq)\right]^{\mathrm{T}}\end{aligned} \tag{25}$$

$$\hat{v}(kq-l) = \hat{y}(kq-l) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq-l)\hat{\boldsymbol{\vartheta}}(kq-q). \tag{26}$$

This DR-HLS algorithm in (22)–(26) requires computing the $n_i \times n_i$ covariance matrix $\boldsymbol{P}_i(kq)$. To initialize the DR-HLS algorithm, we take $\hat{\boldsymbol{\vartheta}}_i(0) = \mathbf{1}_{n_i}/p_0$ and $\boldsymbol{P}_i(0) = p_0 \boldsymbol{I}_{n_i}$.

Next, we compute a- and b-parameters from the estimated α- and β-parameters. Use the estimated parameter vector $\hat{\boldsymbol{\vartheta}}(kq)$ to form the estimates of $\alpha(z^{-1})$ and $\beta(z^{-1})$ as follows:

$$\hat{\alpha}(z^{-1}) = 1 + \hat{\alpha}_1(kq)z^{-q} + \cdots + \hat{\alpha}_{n_a}(kq)z^{-n_a q},$$
$$\hat{\beta}(z^{-1}) = \hat{\beta}_0(kq) + \hat{\beta}_1(kq)z^{-1} + \cdots + \hat{\beta}_m(kq)z^{-m}.$$

Let the estimates of $A(z^{-1})$ and $B(z^{-1})$ be

$$\hat{A}(z^{-1}) = 1 + \hat{a}_1(kq)z^{-1} + \cdots + \hat{a}_{n_a}(kq)z^{-n_a},$$
$$\hat{B}(z^{-1}) = \hat{b}_0(kq) + \hat{b}_1(kq)z^{-1} + \cdots + \hat{b}_{n_b}(kq)z^{-n_b}.$$

According to (3) and (4), let their estimates satisfy the relation $\hat{B}(z^{-1})\hat{\alpha}(z^{-1}) - \hat{A}(z^{-1})\hat{\beta}(z^{-1}) = 0$. Expanding and comparing the coefficients of z^{-i} of both sides establish a set of linear equations of $\hat{a}_i(kq)$ and $\hat{b}_i(kq)$ [16]

$$\boldsymbol{S}(kq)\hat{\boldsymbol{\theta}}(kq) = \boldsymbol{\rho}(kq) \tag{27}$$

where

$$\boldsymbol{S}(kq) = [-\boldsymbol{S}_\beta(kq), \boldsymbol{S}_\alpha(kq)],$$
$$\boldsymbol{\rho}(kq) = \left[\hat{\beta}_0(kq), \hat{\beta}_1(kq), \cdots, \hat{\beta}_m(kq), 0, \cdots, 0\right]^{\mathrm{T}} \in \mathbb{R}^{n_a+m+1}$$

with $\boldsymbol{S}_\alpha(kq) \in \mathbb{R}^{(n_a+m+1)\times(n_b+1)}$ and $\boldsymbol{S}_\beta(kq) \in \mathbb{R}^{(n_a+m+1)\times n_a}$ are matrices

$$\boldsymbol{S}_\alpha(kq) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \gamma_1(kq) & 1 & & \vdots \\ \gamma_2(kq) & \gamma_1(kq) & \ddots & 0 \\ \vdots & \gamma_2(kq) & \ddots & 1 \\ \vdots & \vdots & \ddots & \gamma_1(kq) \\ \gamma_{qn_a}(kq) & \vdots & & \gamma_2(kq) \\ 0 & \gamma_{qn_a}(kq) & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \gamma_{qn_a}(kq) \end{bmatrix},$$

$$\boldsymbol{S}_\beta(kq) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \hat{\beta}_0(kq) & 0 & & \vdots \\ \hat{\beta}_1(kq) & \hat{\beta}_0(kq) & \ddots & 0 \\ \vdots & \hat{\beta}_1(kq) & \ddots & 0 \\ \vdots & \vdots & \ddots & \hat{\beta}_0(kq) \\ \hat{\beta}_m(kq) & \vdots & & \hat{\beta}_1(kq) \\ 0 & \hat{\beta}_m(kq) & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\beta}_m(kq) \end{bmatrix},$$

$$\gamma_l(kq) = \begin{cases} \hat{\alpha}_j(kq), & l = jq, j = 1, 2, \cdots, n_a, \\ 0, & \text{else.} \end{cases}$$

The least squares solution to (27) is given by [30]

$$\hat{\boldsymbol{\theta}}(kq) = \left[\boldsymbol{S}^{\mathrm{T}}(kq)\boldsymbol{S}(kq)\right]^{-1}\boldsymbol{S}^{\mathrm{T}}(kq)\boldsymbol{\rho}(kq). \tag{28}$$

TABLE I
COMPARISON OF COMPUTATIONAL EFFICIENCY

Algorithms	Number of multiplications	Number of additions
RLS	$2n^2+4n$	$2n^2+2n$
DR-HLS	$\sum_{i=1}^{N}(2n_i^2+4n_i)$	$\sum_{i=1}^{N}(2n_i^2+2n_i)$

The following conclusions are made by discussing the RLS and DR-HLS algorithms:

- For system models of high orders, the numbers of the unknown parameters are large, the RLS algorithm suffers computing the covariance matrix $\boldsymbol{P}(kq)$ with high dimensions in each recursion and leads to large computational burden. The DR-HLS algorithm deals with the covariance matrices with smaller dimensions and has a computational efficiency. The numbers of multiplications and additions of the RLS and DR-HLS algorithms for each step are listed in Table I with the flop numbers [29].
- The identification model in (6) can be decomposed into several sub-identification models in (15), and the parameter estimates of the RLS algorithm and the DR-HLS algorithm are convergent but different generally for finite k—see the example later.
- For the DR-HLS algorithm, the number N of the subsystems depends on the computational complexity and estimation accuracy. Under the extreme case with taking $N=n$ and $n_1=n_2=\cdots=n_N=1$, then the DR-HLS algorithm in (22)–(26) reduces to the stochastic gradient (SG) algorithm that has least computational load but lowest estimation accuracy, because the SG algorithm has slower convergence rate than the RLS algorithm for the same data length. Because the computational complexity of the DR-HLS algorithm is related to the ratio q of the input and output samplings, one way is to choose $N=q$.

IV. MAIN CONVERGENCE RESULTS

For notational convenience, the norm of a column vector $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\|^2=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$; the relation $f(k)=o(g(k))$ represents $f(k)/g(k)\to 0$ as $k\to\infty$; for $g(k)\geqslant 0$, we write $f(k)=O(g(k))$ if there exists a positive constant c such that $|f(k)|\leqslant cg(k)$. The symbols $\lambda_{\max}[\boldsymbol{X}]$ and $\lambda_{\min}[\boldsymbol{X}]$ represent the maximum and minimum eigenvalues of the symmetric matrix $\boldsymbol{X}$, respectively.

Assume that $\{v(k)\}$ is a martingale difference sequence defined on a probability space $(\Omega,\mathcal{F},P)$, where $\{\mathcal{F}_k\}$ is the σ algebra sequence generated by $\{v(k)\}$, i.e., $\mathcal{F}_k=\sigma(v(k),v(k-1),v(k-2),\cdots,)$. The sequence $\{v(k)\}$ satisfies the noise assumptions [28]

$$\text{(A1)}\quad \mathrm{E}\left[v(k)|\mathcal{F}_{k-1}\right]=0,\ \text{a.s.}$$
$$\text{(A2)}\quad \mathrm{E}\left[v^2(k)|\mathcal{F}_{k-1}\right]\leqslant\sigma^2<\infty,\ \text{a.s.}$$

Defining $r_i(kq)=\mathrm{tr}[\boldsymbol{P}_i^{-1}(kq)]$, it follows that $r_i(kq)\leqslant n_i\lambda_{\max}[\boldsymbol{P}_i^{-1}(kq)]$, and

$$\ln\left|\boldsymbol{P}_i^{-1}(kq)\right|=O\left(\ln r_i(kq)\right)=O\left(\ln\lambda_{\max}\left[\boldsymbol{P}_i^{-1}(kq)\right]\right). \tag{29}$$

Theorem 1: For the system in (6) and the DR-HLS algorithm in (22)–(26), assume that (A1) and (A2) hold, then

$$\left\|\tilde{\boldsymbol{\vartheta}}_i(kq)\right\|^2=O\left(\frac{[\ln r_i(kq)]^c}{\lambda_{\min}\left[\boldsymbol{P}_i^{-1}(kq)\right]}\right),\ \text{a.s., for any } c>1.$$

Theorem 1 shows that for the noise sequence $\{v(k)\}$ with a bounded variance, the convergent rate of the DR-HLS parameter estimates is the ratio of the logarithm of the maximum eigenvalue to the minimum eigenvalue of the covariance matrix $\boldsymbol{P}_i^{-1}(kq)$. The proof is given in the Appendix by using a similar way in [31].

TABLE II
THE DR-HLS ESTIMATES OF $\boldsymbol{\theta}$

k	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	δ (%)
100	-0.96850	0.55218	-0.56861	0.60133	-0.38264	0.60026	20.33981
200	-1.20513	0.73981	-0.49615	0.55010	-0.46153	0.54757	21.91827
500	-1.17140	0.56427	-0.35648	0.50622	-0.38525	0.47526	7.72114
1000	-1.11948	0.55221	-0.39294	0.48896	-0.35441	0.48145	4.32078
2000	-1.10747	0.55541	-0.40611	0.47666	-0.36244	0.51671	4.53002
3000	-1.08336	0.54047	-0.41981	0.46961	-0.34245	0.51377	4.26584
4000	-1.10290	0.56091	-0.42248	0.46482	-0.34262	0.51409	5.05574
5000	-1.10184	0.53984	-0.40210	0.46022	-0.33896	0.50821	3.02059
6000	-1.09190	0.52360	-0.39479	0.46110	-0.33765	0.50924	2.09372
$\boldsymbol{\theta}$	-1.10000	0.51000	-0.37000	0.46000	-0.34000	0.50000	

TABLE III
THE RLS ESTIMATES OF $\boldsymbol{\theta}$

k	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	δ (%)
100	-1.07808	0.64377	-0.54190	0.52474	-0.39636	0.60498	17.42411
200	-1.20299	0.66793	-0.42474	0.51193	-0.42167	0.50188	14.81157
500	-1.19233	0.57450	-0.34566	0.49296	-0.37845	0.46272	8.88210
1000	-1.14653	0.55622	-0.36886	0.48024	-0.35765	0.47289	5.13433
2000	-1.12777	0.57259	-0.40163	0.47147	-0.36655	0.51304	5.53681
3000	-1.10212	0.55949	-0.41975	0.46526	-0.34654	0.51065	4.83830
4000	-1.10892	0.56726	-0.42209	0.46218	-0.34273	0.51224	5.34221
5000	-1.10564	0.54524	-0.40256	0.45809	-0.33883	0.50730	3.30931
6000	-1.09522	0.53069	-0.39790	0.45928	-0.33745	0.50816	2.44248
$\boldsymbol{\theta}$	-1.10000	0.51000	-0.37000	0.46000	-0.34000	0.50000	

V. EXAMPLE

Consider a third-order system with

$$\begin{aligned}A(z^{-1})&=1+a_1z^{-1}+a_2z^{-2}+a_3z^{-3}\\&=1-1.10z^{-1}+0.50z^{-2}-0.37z^{-3},\\B(z^{-1})&=b_1z^{-1}+b_2z^{-2}+b_3z^{-3}\\&=0.46z^{-1}-0.34z^{-2}+0.50z^{-3}.\end{aligned}$$

Taking $q=2$, and the transformation polynomial $D(z^{-1})=1-a_1z^{-1}+a_2z^{-2}-a_3z^{-3}$, then the corresponding dual-rate model may be expressed as

$$\begin{aligned}\alpha(z^{-1})y(k)&=\beta(z^{-1})u(k)+D(z^{-1})v(k),\\\alpha(z^{-1})&=1-0.1900z^{-2}-0.5539z^{-4}-0.1369z^{-6},\\\beta(z^{-1})&=0.4600z^{-1}+0.1660z^{-2}+0.3606z^{-3}\\&\quad+0.5468z^{-4}+0.1292z^{-5}+0.1850z^{-6},\\\boldsymbol{\vartheta}&=[-0.1900,-0.5539,-0.1369,0.4600,0.1660,\\&\quad 0.3606,0.5468,0.1292,0.1850]^{\mathrm{T}}.\end{aligned}$$

Here $\{u(k)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance, $\{v(k)\}$ is independent of $\{u(k)\}$ and is taken as a white noise sequence with zero mean and variance $\sigma^2=0.25$, the corresponding noise-to-signal ratio is $\delta_{\mathrm{ns}}=82.80\%$ which is defined by the square root of the ratio of the variances of $w(k)$ and $x(k)$ in Fig. 1, i.e.

$$\delta_{\mathrm{ns}}=\sqrt{\frac{\mathrm{var}\left[w(k)\right]}{\mathrm{var}\left[x(k)\right]}}\times 100\%.$$

We decompose this example system into $N=2$ subsystems. Applying the DR-HLS algorithm and the RLS algorithm to estimate the dual-rate model parameters $\boldsymbol{\vartheta}$ and using (27), (28) to compute the parameter estimates of a_i and b_i, the results are as shown in Tables II, III, where the estimation errors $\delta=\|\hat{\boldsymbol{\theta}}(k)-\boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ versus k are shown in Fig. 2.

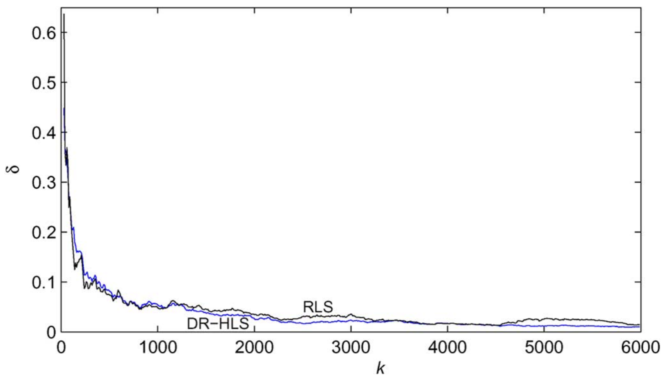

Fig. 2. Parameter estimation errors δ versus k $(N = 2)$.

In the simulation, the example model is decomposed into $N = 2$ sub-identification models with $n_1 = 4$ and $n_2 = 5$, the numbers of multiplications and additions of the RLS algorithm in each recursion are 198 and 180, respectively, while those of the DR-HLS algorithm are 118 and 100. Thus the RLS algorithm has 378 flops while the DR-HLS algorithm has 218 flops. This indicates that the DR-HLS algorithm conducts a large reduction of computation effort compared with the RLS algorithm.

From Tables II, III and Fig. 2, the identification accuracy of the DR-HLS algorithm is very close to that of the RLS algorithm for $N = 2$, but the proposed algorithm has much less computational efforts.

VI. Conclusion

In this technical note, we introduced the polynomial transformation technique and proposed a hierarchical least squares identification algorithm for dual-rate linear systems with noises. The proposed algorithm offers accurate parameter estimation for the dual-rate ARMAX models with moving average noise and requires significantly less computational effort than the RLS algorithm and the convergence properties of the DR-HLS algorithm were analyzed. The simulation results have confirmed the effectiveness of the proposed method. The developed method can combine the multivation identification methods [32]–[36] to identify non-uniformly sampled systems [37], [38], multivariable systems [39], nonlinear systems [40], and missing-data systems or scarce measurement systems [41], [42].

Appendix
The Proofs of Lemmas and Theorem

The following lemmas are required to establish the main convergence results.

Lemma 1: For the algorithm in (22)–(26), the following inequality holds

$$\sum_{i=1}^{\infty} \frac{\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(iq)\boldsymbol{P}_i(iq)\hat{\boldsymbol{\phi}}_i(iq)}{\left[\ln\left|\boldsymbol{P}_i^{-1}(iq)\right|\right]^c} < \infty, \text{ a.s., for any } c > 1.$$

The proof can be done in a similar way to that of Lemma 1 in [16], [31].

Lemma 2: For the system in (6) and the algorithm in (22)–(26), assume that (A1) and (A2) hold, and

$$\text{(A3)} \quad H(z^{-1}) = ND^{-1}(z^{-1}) - \frac{1}{2}$$

is strictly positive real, then

$$\begin{aligned}\mathrm{E}\left[V_i(kq)|\mathcal{F}_{kq-q}\right] &\leqslant V_i(kq-q) + S_i(kq-q)\\ &\quad +2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\sigma^2, \text{ a.s.}\end{aligned} \quad (30)$$

Here, (A3) guarantees that $S_i(kq) \geqslant 0$.

Proof: Define

$$\begin{aligned}e(kq) &= y(kq) - \hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}(kq-q)\\ &= y(kq) - \sum_{i=1}^{N}\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}_i(kq-q)\\ &= \hat{v}(kq) + \sum_{i=1}^{N}\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)e(kq)\\ &= \frac{\hat{v}(kq)}{1-\sum_{i=1}^{N}\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)}.\end{aligned}$$

Define

$$\tilde{\boldsymbol{\vartheta}}_i(kq) = \hat{\boldsymbol{\vartheta}}_i(kq) - \boldsymbol{\vartheta}_i = \tilde{\boldsymbol{\vartheta}}_i(kq-q) + \boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)e(kq). \quad (31)$$

Thus we have

$$\begin{aligned}&\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq-q)\tilde{\boldsymbol{\vartheta}}_i(kq)\\ &\quad= \tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq-q)\tilde{\boldsymbol{\vartheta}}_i(kq-q)\\ &\qquad+ \tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq-q)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)e(kq)\\ &\quad= \tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq-q)\boldsymbol{P}_i^{-1}(kq-q)\tilde{\boldsymbol{\vartheta}}_i(kq-q)\\ &\qquad+ \hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\boldsymbol{P}_i^{-1}(kq-q)\\ &\qquad\times\left[\tilde{\boldsymbol{\vartheta}}_i(kq) - \boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)e(kq)\right]e(kq)\\ &\qquad+ \tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq-q)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)e(kq).\end{aligned} \quad (32)$$

From (24), we have $\boldsymbol{P}_i^{-1}(kq) = \boldsymbol{P}_i^{-1}(kq-q) + \hat{\boldsymbol{\phi}}_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)$. Define

$$\begin{aligned}\mu_i(kq) &= \frac{1-\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)}{1-\sum_{i=1}^{N}\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)},\\ \tilde{y}_i(kq) &= \frac{1}{2}\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\phi}}_i(kq) + \mu_i(kq)\left[\hat{v}(kq) - v(kq)\right],\\ \tilde{u}_i(kq) &= -\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\phi}}_i(kq), \quad S_i(kq) = 2\sum_{j=1}^{k}\tilde{u}_i(jq)\tilde{y}_i(jq),\\ V_i(kq) &= \tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq).\end{aligned}$$

Using (31), it follows that:

$$\begin{aligned}V_i(kq) &= \tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq-q)\tilde{\boldsymbol{\vartheta}}_i(kq) + \left[\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\right]^2\\ &= V_i(kq-q) + \left[\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\right]^2\\ &\quad + 2\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\phi}}_i(kq)e(kq) - \hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\\ &\quad \times\left[1-\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\right]e^2(kq)\\ &\quad - 2\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)e(kq)\\ &\leqslant V_i(kq-q) + \left[\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\right]^2 + 2\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\\ &\quad \times\left[1-\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\right]e(kq)\\ &= V_i(kq-q) + 2\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\left\{\frac{1}{2}\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\right.\\ &\qquad \left.+\left[1-\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\right]e(kq)\right\}\\ &= V_i(kq-q) - 2\tilde{u}_i(kq)\tilde{y}_i(kq)\\ &\quad + 2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq-q)v(kq)\\ &\quad + 2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)\\ &\quad \times\left\{\left[e(kq)-v(kq)\right]v(kq) + v^2(kq)\right\}.\end{aligned}$$

Adding $S_i(kq)$ to both sides gives

$$\begin{aligned} V_i(kq)+S_i(kq) \leqslant & V_i(kq-q)+S_i(kq-q) \\ & +2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq-q)v(kq) \\ & +2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq) \\ & \times\left\{[e(kq)-v(kq)]\,v(kq)+v^2(kq)\right\}. \end{aligned}$$

Since $V_i(kq-q)$, $S_i(kq-q)$, $\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\phi}}_i(kq)$ and $\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)[e(kq)-v(kq)]$ are uncorrelated with $v(kq)$ and $\mathcal{F}_{kq-q}$ measurable, taking the conditional expectation with respect to $\mathcal{F}_{kq-q}$ and using (A1) and (A2) lead to (30). Next, we show $S_i(kq)\geqslant 0$. Since

$$\begin{aligned} & D(z^{-1})\,[\hat{v}(kq)-v(kq)] \\ &= D(z^{-1})\hat{v}(kq)-\alpha(z^{-1})y(kq)+\beta(z^{-1})u(kq) \\ &= \hat{v}(kq)-y(kq)+\hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\boldsymbol{\vartheta}=-\hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\hat{\boldsymbol{\vartheta}}(kq)+\hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\boldsymbol{\vartheta} \\ &= -\hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\left[\hat{\boldsymbol{\vartheta}}(kq)-\boldsymbol{\vartheta}\right]=-\hat{\boldsymbol{\phi}}^{\mathrm{T}}(kq)\tilde{\boldsymbol{\vartheta}}(kq)=\tilde{u}(kq) \end{aligned} \tag{33}$$

we have

$$\begin{aligned} \tilde{y}(kq) &= \sum_{i=1}^{N}\left\{\frac{1}{2}\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\hat{\boldsymbol{\phi}}_i(kq)+\mu_i(kq)\,[\hat{v}(kq)-v(kq)]\right\} \\ &= -\frac{1}{2}\sum_{i=1}^{N}\tilde{u}_i(kq)+\sum_{i=1}^{N}\mu_i(kq)D^{-1}(z^{-1})\tilde{u}(kq) \\ &= -\frac{1}{2}\tilde{u}(kq)+ND^{-1}(z^{-1})\tilde{u}(kq) \\ &\quad \times\left[ND^{-1}(z^{-1})-\frac{1}{2}\right]\tilde{u}(kq) \\ &= H(z^{-1})\tilde{u}(kq). \end{aligned}$$

Since $\tilde{y}(kq)$ is the output of the linear system $H(z^{-1})$ driven by $\tilde{u}(kq)$ and $H(z^{-1})$ is strictly positive real, we have $S_i(kq)\geqslant 0$ according to Appendix C in [28]. This proves Lemma 2. □

Proof of Theorem 1: Let

$$W_i(kq)=\frac{V_i(kq)+S_i(kq)}{\left[\ln\left|\boldsymbol{P}_i^{-1}(kq)\right|\right]^c},\quad c>1.$$

Since $\ln|\boldsymbol{P}_i^{-1}(kq)|$ is nondecreasing, using Lemma 2 gives

$$\begin{aligned} & \mathrm{E}\,[W_i(kq)|\mathcal{F}_{kq-q}] \\ & \leqslant \frac{V_i(kq-q)+S_i(kq-q)}{\left[\ln\left|\boldsymbol{P}_i^{-1}(kq)\right|\right]^c}+\frac{2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)}{\left[\ln\left|\boldsymbol{P}_i^{-1}(kq)\right|\right]^c} \\ & \leqslant W_i(kq-q)+\frac{2\mu_i(kq)\hat{\boldsymbol{\phi}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i(kq)\hat{\boldsymbol{\phi}}_i(kq)}{\left[\ln\left|\boldsymbol{P}_i^{-1}(kq)\right|\right]^c},\ \text{a.s.} \end{aligned}$$

Applying the martingale convergence theorem [28], we have

$$W_i(kq)=\frac{V_i(kq)+S_i(kq)}{\left[\ln\left|\boldsymbol{P}_i^{-1}(kq)\right|\right]^c}\to C<\infty,\ \text{a.s.}$$

This means $V_i(kq)=O([\ln|\boldsymbol{P}_i^{-1}(kq)|]^c)$, a.s. Using (29) gives $V_i(kq)=O([\ln r_i(kq)]^c)$, a.s. Thus, we have

$$\left\|\tilde{\boldsymbol{\vartheta}}_i(kq)\right\|^2\leqslant\frac{\mathrm{tr}\left[\tilde{\boldsymbol{\vartheta}}_i^{\mathrm{T}}(kq)\boldsymbol{P}_i^{-1}(kq)\tilde{\boldsymbol{\vartheta}}_i(kq)\right]}{\lambda_{\min}\left[\boldsymbol{P}_i^{-1}(kq)\right]}=\frac{V_i(kq)}{\lambda_{\min}\left[\boldsymbol{P}_i^{-1}(kq)\right]}.$$

This gives the conclusion of Theorem 1. □

REFERENCES

[1] S. Albert, H. Hiden, A. Conlin, E. B. Martin, G. A. Montague, and A. J. Morris, "Inferential quality assessment in breakfast cereal production," *J. Food Eng.*, vol. 50, no. 3, pp. 157–166, 2001.

[2] N. Lu, Y. Yang, F. Cao, and F. Wang, "Multirate dynamic inferential modeling for multivariable processes," *Chem. Eng. Sci.*, vol. 59, no. 4, pp. 855–864, 2004.

[3] J. H. Ahrens, X. Tan, and H. K. Khalil, "Multirate sampled-data output feedback control with application to smart material actuated systems," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2518–2529, Nov. 2009.

[4] J. Wang, T. Chen, and B. Huang, "Multirate sampled-data systems: Computing fast-rate models," *J. Process Control*, vol. 14, no. 1, pp. 79–88, 2004.

[5] W. Li, Z. Han, and S. L. Shah, "Subspace identification for FDI in systems with non-uniformly sampled multirate data," *Automatica*, vol. 42, no. 4, pp. 619–627, 2006.

[6] W. Li, S. L. Shah, and D. Y. Xiao, "Kalman filters in non-uniformly sampled multirate systems: For FDI and beyond," *Automatica*, vol. 44, no. 1, pp. 199–208, 2008.

[7] X. Wang, B. Huang, and T. Chen, "Multirate minimum variance control design and control performance assessment: A data-driven subspace approach," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 1, pp. 65–74, Jan. 2007.

[8] B. Yu, Y. Shi, and H. Huang, " L_2-L-infinity filtering for multirate systems using lifted models," *Circuits, Syst., Signal Processing*, vol. 27, no. 5, pp. 699–711, 2008.

[9] S. Mo, X. Chen, J. Zhao, J. Qian, and Z. Shao, "A two-stage method for identification of dual-rate systems with fast input and very slow output," *Ind. Eng. Chem. Res.*, vol. 48, no. 4, pp. 1980–1988, 2008.

[10] X. Liu, H. J. Marquez, and Y. Lin, "Input-to-state stabilization for nonlinear dual-rate sampled-data systems via approximate discrete-time model," *Automatica*, vol. 44, no. 12, pp. 3157–3161, 2008.

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

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

[13] F. Ding and T. Chen, "Hierarchical identification of lifted state-space models for general dual-rate systems," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 6, pp. 1179–1187, Jun. 2005.

[14] P. Qin, S. Kanae, Z. Yang, and K. Wada, "Identification of lifted models for general dual-rate sampled-data systems using N4SID algorithm," *IEEJ Trans. Electron., Inform. Syst.*, vol. 128, no. 5, pp. 788–794, 2008.

[15] W. P. Lu and G. Fisher, "Output estimation with multi-rate sampling," *Int. J. Control*, vol. 48, no. 1, pp. 149–160, 1988.

[16] F. Ding, P. X. Liu, and Y. Shi, "Convergence analysis of estimation algorithms of dual-rate stochastic systems," *Appl. Math. Comput.*, vol. 176, no. 1, pp. 245–261, 2006.

[17] J. Ding, Y. Shi, H. Wang, and F. Ding, "A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled-data systems," *Digital Signal Processing*, vol. 20, no. 4, pp. 1238–1247, 2010.

[18] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.

[19] R. B. Gopaluni, "Identification of nonlinear processes with known structure under missing observations," in *Proc. 17th IFAC World Congress*, Seoul, Korea, Jul. 6–11, 2008, pp. 6478–6483.

[20] R. B. Gopaluni, "A particle filter approach to identification of nonlinear processes under missing observations," *Canadian J. Chem. Eng.*, vol. 86, no. 6, pp. 1081–1092, 2008.

[21] A. J. Isaksson, "Identification of ARX-models subject to missing data," *IEEE Trans. Autom. Control*, vol. 38, no. 5, pp. 813–819, May 1993.
[22] J. Ding, F. Ding, and P. X. Liu, "Hierarchical least squares parameter estimation algorithms for dual-rate sampled-data systems," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, Victoria, BC, Canada, May 12–15, 2008, pp. 536–541.
[23] G. Box and G. Jenkins, *Time Series Analysis-Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.
[24] F. Ding, P. X. Liu, and H. Z. Yang, "Parameter identification and intersample output estimation for dual-rate systems," *IEEE Trans. Syst., Man, Cybern. A*, vol. 38, no. 4, pp. 966–975, Jul. 2008.
[25] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
[26] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. New York: Wiley, 1998.
[27] F. Ding and T. Chen, "Hierarchical gradient-based identification of multivariable discrete-time systems," *Automatica*, vol. 41, no. 2, pp. 315–325, 2005.
[28] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[30] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
[31] 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.
[32] F. Ding and T. Chen, "Performance analysis of multi-innovation gradient type identification methods," *Automatica*, vol. 43, no. 1, pp. 1–14, 2007.
[33] Y. J. Liu, Y. S. Xiao, and X. L. Zhao, "Multi-innovation stochastic gradient algorithm for multiple-input single-output systems using the auxiliary model," *Appl. Math. Comput.*, vol. 215, no. 4, pp. 1477–1483, 2009.
[34] L. L. Han and F. Ding, "Multi-innovation stochastic gradient algorithms for multi-input multi-output systems," *Digital Signal Processing*, vol. 19, no. 4, pp. 545–554, 2009.
[35] F. Ding, "Several multi-innovation identification methods," *Digital Signal Processing*, vol. 20, no. 4, pp. 1027–1039, 2010.
[36] D. Q. Wang and F. Ding, "Performance analysis of the auxiliary models based multi-innovation stochastic gradient estimation algorithm for output error systems," *Digital Signal Processing*, vol. 20, no. 3, pp. 750–762, 2010.
[37] F. Ding, L. Qiu, and T. Chen, "Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems," *Automatica*, vol. 45, no. 2, pp. 324–332, 2009.
[38] F. Ding, G. Liu, and X. P. Liu, "Partially coupled stochastic gradient identification methods for non-uniformly sampled systems," *IEEE Trans. Autom. Control*, vol. 55, no. 8, pp. 1976–1981, Aug. 2010.
[39] Y. J. Liu, J. Sheng, and R. F. Ding, "Convergence of stochastic gradient estimation algorithm for multivariable ARX-like systems," *Comp. Math. Appl.*, vol. 59, no. 8, pp. 2615–2627, 2010.
[40] F. Ding, P. X. Liu, and G. Liu, "Identification methods for Hammerstein nonlinear systems," *Digital Signal Processing*, vol. 21, no. 2, pp. 215–238, 2011.
[41] F. Ding and J. Ding, "Least squares parameter estimation with irregularly missing data," *Int. J. Adaptive Control Signal Processing*, vol. 24, no. 7, pp. 540–553, 2010.
[42] F. Ding, G. Liu, and X.P. Liu, "Parameter estimation with scarce measurements," *Automatica*, vol. 47, no. 8, pp. 1646–1655, 2011.

Stability-Preserving Optimization in the Presence of Fast Disturbances

Benedikt Wirth, Johannes Gerhard, and Wolfgang Marquardt

***Abstract*—We present algebraic conditions on the trajectory of a dynamical system to approximately describe a certain type of system robustness. The corresponding equations can be used as constraints in a robust optimization procedure to select a set of optimal design parameters for a dynamical system which is subject to fast disturbances. Robustness is ensured by requiring the disturbance parameters to stay sufficiently far away from critical manifolds in the disturbance parameter space, at which the system would lose stability. The closest distance to the critical manifolds is measured along their normal vectors.**

***Index Terms*—Dynamical systems, fast disturbances, robust optimization, robust stability.**

I. Introduction

Optimization of design parameters for dynamical systems represents an important task in engineering. Usually, the economic profit of a system is maximized under some feasibility and stability conditions. Most often, the objective function of this optimization can be expressed as a function $\phi(x,p)$ of the chosen design parameters $p \in \mathbb{R}^{n_p}$ and the system state $x \in \mathbb{R}^{n_x}$ which is governed by the dynamic system

$$\dot{x} = f(x, p, \delta(d,t)), \quad x(t_0) = x_0 \tag{1}$$

with $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \times \mathbb{R}^{n_\delta} \to \mathbb{R}^{n_x}$. The dot refers to the time derivative, and $t \in \mathbb{R}$ denotes time. We assume a smooth solution of (1) to exist for arbitrary initial conditions x_0 and any time t. $\delta(d,t)$ represents a disturbance acting on the system, which is parameterized by the disturbance parameters $d \in \mathbb{R}^{n_d}$. More precisely, we introduce the following

Definition 1 (Disturbance): Let $\delta(d,\cdot) : \mathbb{R} \to \mathbb{R}^{n_\delta}$ be a given family of functions, parametrized over $d \in \mathbb{R}^{n_d}$, such that $\delta(d,\cdot)$ is continuous on $(0,\infty)$ and $\delta(d,\cdot) \equiv 0$ on $(-\infty, 0]$. A dynamical system (1) is then called a *disturbed system*. A *disturbance* is a particular instance $\delta(\hat{d},\cdot)$, $\hat{d} \in \mathbb{R}^{n_d}$, of this family. A *step disturbance* is a disturbance $\delta(\hat{d},\cdot)$ with $\delta(\hat{d},t) = \text{const.}$ for $t > 0$.

In different words, in a disturbed system, $\delta(d,t)$ suddenly changes from zero to some (possibly time-varying) function at $t = 0$. If this change happens on a time-scale faster than the system dynamics such that a quasi-steady approximation is meaningless, one speaks of fast disturbances [1]. In particular, we will in the following restrict ourselves to the case of step disturbances. Since we can identify each disturbance $\delta(d,\cdot)$ with its disturbance parameter d, for simplicity we will

Manuscript received February 01, 2009; revised November 26, 2010; accepted May 26, 2011. Date of publication June 20, 2011; date of current version November 02, 2011. This work was supported by Deutsche Forschungsgemeinschaft (DFG) under grant MA1188/22-1. Recommended by Associate Editor H. Ito.

B. Wirth is with the AVT—Process Systems Engineering, RWTH Aachen University, Aachen, Germany and also with the Courant Institute, New York University, New York, NY 10012 USA (e-mail: benedikt.wirth@ins.uni-bonn.de).

J. Gerhard is with the AVT—Process Systems Engineering, RWTH Aachen University, Aachen, Germany and also with the Evonik Degussa GmbH, Essen 45130, Germany (e-mail: johannes.gerhard@web.de).

W. Marquardt is with the AVT—Process Systems Engineering, RWTH Aachen University, Aachen 52056, Germany (e-mail: wolfgang.marquardt@avt.rwth-aachen.de).

Digital Object Identifier 10.1109/TAC.2011.2160029