Least squares estimation for a class of non-uniformly sampled systems based on the hierarchical identification principle

Yanjun Liu · Feng Ding · Yang Shi

Abstract This paper presents a novel hierarchical least squares algorithm for a class of non-uniformly sampled systems. Based on the hierarchical identification principle, the identification model with a high dimensional parameter vector is decomposed into a group of submodels with lower dimensional parameter vectors. By using the least squares method to identify the submodels and taking a coordinated measure to address the associated items between the submodels, all the system parameters can be estimated. The proposed algorithm can save the computation cost. The performance analysis indicates that parameter estimates converge to their true values. The simulation tests confirm the convergence results.

1 Introduction

Conventional discrete-time control systems are mainly developed for single-rate systems in which all input and output variables are sampled at a single (uniform) rate [2]. Identification techniques for single-rate systems have been well investigated [1, 16,

Y. Liu (✉) · F. Ding
Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, Ministry of Education, Wuxi 214122, P.R. China
e-mail: yanjunliu_1983@126.com

F. Ding
e-mail: fding@jiangnan.edu.cn

Y. Shi
Department of Mechanical Engineering, University of Victoria, Victoria, BC V8N 3P6, Canada
e-mail: yshi@uvic.ca

25, 26]. However, the single-rate sampling scheme sometimes may not be applicable in practical industrial processes, because the involved physical variables might be sampled at different sampling rates; see some examples in [13, 20]. Such systems with more than one sampling rate are called multirate systems. Recently, multirate systems have been a hot topic in control and identification fields, and many related achievements have been reported [3, 4, 15, 18, 23, 33, 39].

Most of the sampling schemes for multirate systems in the literature are uniform sampling, that is, the sampling interval for all variables are fixed. However, a more general class of multirate systems feature the following characteristics: The involved variables are measured at non-uniform intervals, e.g., in the cases when manual sampling or laboratory analysis is required. Systems whose sampling intervals for the input and/or output channels are non-equidistant in time are termed as non-uniformly sampled multirate systems [8, 21, 24, 28]. The systems with irregular missing samples can be characterized under the framework of the non-uniformly sampled system model [10, 14, 17, 31, 32, 38, 40, 41, 43, 45]. In this paper, we consider a class of periodically non-uniformly sampled systems which are special cases of the general non-uniformly sampled systems. It has been shown that the periodically non-uniform sampling has some promising properties, such as preserving controllability and observability [8, 28] and being capable of uniquely recovering the continuous-time system [8]. The related identification and control problems have been investigated for periodically non-uniformly sampled systems [8, 9, 21, 22, 24, 27, 28]. Specifically, some interesting identification methods have been proposed for the non-uniformly sampled systems, for example, the subspace methods [21, 27], the auxiliary model-based least squares method [24], the multi-innovation gradient method [37] and the partially coupled least squares method [9]. In this work, the main objective is to propose a computationally efficient identification algorithm for periodically non-uniformly sampled systems based on a hierarchical identification principle.

It is known that multirate and non-uniformly sampled systems involve more parameters than single-rate systems, which will unavoidably result in heavy computational costs of the identification algorithms. This motivates us to develop computationally efficient algorithms for online identification, which is of both theoretical merit and practical needs. The decomposition technique is useful in many areas such as the image procession [34–36]. Based on the decomposition technique, the hierarchical identification principle is very effective in reducing the complexity of the identification algorithm for large-scale systems [5]. The key is to decompose a system into subsystems and to address the associated items [6, 11]. For system identification models with different structures, the ways of doing decomposition are different. The multivariable system in [5] was decomposed into two subsystems, one with a parameter vector and the other with a parameter matrix. The dual-rate state-space model in [6] was also decomposed into two subsystems according to the lifted state-space structure. In this paper, we propose to decompose the system identification model into N ($2 \leqslant N \leqslant \dim(\boldsymbol{\theta})$) submodels, where $\dim(\boldsymbol{\theta})$ represents the dimension of the parameter vector $\boldsymbol{\theta}$. Based on such a decomposition, the proposed algorithm can greatly reduce the computational complexity.

The rest of the paper is organized as follows. Section 2 derives the identification model of the non-uniformly sampled systems and describes the problem formulation.

Section 3 gives the recursive least squares algorithm for the non-uniformly sampled systems for comparisons. Section 4 presents a hierarchical least squares algorithm and compares the computational load with the recursive least squares algorithm. Section 5 analyzes the performance of the proposed algorithm. Section 6 provides an illustration example. Finally, we offer some concluding remarks in Sect. 7.

2 Model description and problem formulation

Consider the systems with a periodically non-uniform updating and uniform sampling scheme, which can refer to the Fig. 3 in [24]. For such a non-uniformly sampled system, the control input is non-uniformly updated r times with intervals τ_i ($i = 1, 2, \ldots, r$) at the time instants $t = kT + t_i$ $i = 0, 1, 2, \ldots, r - 1$. Here, $t_i := t_{i-1} + \tau_i = \tau_1 + \tau_2 + \cdots + \tau_i$, $t_0 = 0$ and $T := \tau_1 + \tau_2 + \cdots + \tau_r = t_r$ is known as the frame period; the output is sampled uniformly with the frame period T. In the following, we establish the discrete-time state-space model and the input–output representation of the non-uniformly sampled systems from its continuous-time state-space model.

Consider a continuous-time process with the controllable and observable state-space representation:

$$\begin{cases} \dot{x}(t) = A_c x(t) + B_c u(t), \\ y(t) = C x(t) + D u(t), \end{cases} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}$ the control input, $y(t) \in \mathbb{R}$ the output, A_c, B_c, C are constant matrices of appropriate dimensions and D is a constant. Because of the non-uniform zero-order hold at the input port, the input $u(t)$ is a square wave signal with the following expression:

$$u(t) = \begin{cases} u(kT) & kT \leqslant t < kT + t_1, \\ u(kT + t_1) & kT + t_1 \leqslant t < kT + t_2, \\ \vdots \\ u(kT + t_{r-1}) & kT + t_{r-1} \leqslant t < (k+1)T. \end{cases} \tag{2}$$

The output $y(t)$ is sampled by a sampler with the frame period T, yielding a discrete-time signal $y(kT)$.

The solution for the state equation in (1) is given by

$$x(t) = e^{A_c(t-t_0)} x(t_0) + \int_{t_0}^{t} e^{A_c(t-\tau)} B_c u(\tau) \, d\tau. \tag{3}$$

For a non-pathological frame period T, by letting $t_0 = kT$ and $t = kT + T$ in (3) and using (2) to discretize the state-space model in (1), we have [8, 24],

$$\begin{cases} x(kT + T) = A x(kT) + \sum_{i=1}^{r} B_i u(kT + t_{i-1}), \\ y(kT) = C x(kT) + D u(kT), \end{cases} \tag{4}$$

where

$$A := \mathrm{e}^{A_c T} \in \mathbb{R}^{n \times n}, \qquad B_i := \mathrm{e}^{A_c(T-t_i)} \int_0^{\tau_i} \mathrm{e}^{A_c t}\, \mathrm{d}t\, B_c \in \mathbb{R}^n.$$

Transforming the state-space model in (4) into an input–output representation and taking into account the disturbance $v(kT)$, we have

$$\alpha(z)y(kT) = \sum_{i=1}^r b_i(z)u(kT + t_{i-1}) + v(kT), \qquad (5)$$

where $\{v(kT)\}$ is a white noise sequence with zero mean, $a(z)$ and $b_i(z)$ are polynomials, in the unit backward shift operator z^{-1}, related to A, B_i, C and D and have the form of [24]

$$a(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}, \quad a_i \in \mathbb{R}^1,$$
$$b_1(z) := b_{10} + b_{11} z^{-1} + b_{12} z^{-2} + \cdots + b_{1n} z^{-n}, \quad b_{1j} \in \mathbb{R}^1,$$
$$b_i(z) := b_{i1} z^{-1} + b_{i2} z^{-2} + \cdots + b_{in} z^{-n}, \quad b_{ij} \in \mathbb{R}^1, \ i = 2, 3, \ldots, r.$$

Remark 1 Equation (5) is the input–output representation for the non-uniformly sampled systems. In the following, we will develop the identification algorithm for the model in (5). $\{u(kT + t_i), y(kT) : i = 0, 1, 2, \ldots, r-1; k = 1, 2, \ldots\}$ are the available input–output data. a_i and b_{ij} are the parameters to be identified. In [24], an auxiliary model-based least squares method has been developed for the non-uniformly sampled systems with an output error model. In this paper, we aim to develop a new method to reduce the computational burden of the least squares algorithm.

Remark 2 The expression in (5) can be viewed as a multiple-input single-output system model with r fictitious inputs $u(kT + t_i)$, $i = 0, 1, 2, \ldots, r-1$. When $r = 1$, it becomes a discrete-time model for single-rate systems. It can be seen from (5) that the number of parameters for the non-uniformly sampled model is $(r + 1)n + 1$, which is much larger than $2n + 1$ of the single-rate model. Thus the computational burden of the corresponding identification algorithm for the non-uniformly sampled model is heavier than that for the single-rate model.

3 The recursive least squares algorithm

In order to show the advantages of the proposed algorithm in this paper, we recall the well-established recursive least squares (RLS) algorithm in this section.

Let the superscript T denote the vector transpose and define the parameter vector $\boldsymbol{\theta}$ and the information vector $\boldsymbol{\varphi}(kT)$ as

$$\boldsymbol{\theta} := [a_1, a_2, \ldots, a_n, b_{10}, b_{11}, b_{12}, \ldots, b_{1n}, b_{21}, b_{22}, \ldots, b_{2n}, \ldots,$$
$$b_{r1}, b_{r2}, \ldots, b_{rn}]^{\mathrm{T}} \in \mathbb{R}^{n_0}, \quad n_0 := (r+1)n + 1,$$
$$\boldsymbol{\varphi}(kT) := \big[-y(kT - T), -y(kT - 2T), \ldots, -y(kT - nT), u(kT), u(kT - T),$$
$$u(kT - 2T), \ldots, u(kT - nT), u(kT + t_1 - T), u(kT + t_1 - 2T), \ldots,$$

$$u(kT + t_1 - nT), \ldots, u(kT + t_{r-1} - T), u(kT + t_{r-1} - 2T), \ldots,$$

$$u(kT + t_{r-1} - nT)\big]^{\mathrm{T}} \in \mathbb{R}^{n_0}.$$

Equation (5) can be written in a vector form

$$y(kT) = \boldsymbol{\varphi}^{\mathrm{T}}(kT)\boldsymbol{\theta} + v(kT). \tag{6}$$

The estimate $\hat{\boldsymbol{\theta}}(kT)$ of the parameter vector $\boldsymbol{\theta}$ in (6) can be obtained from the following RLS algorithm:

$$\hat{\boldsymbol{\theta}}(kT) = \hat{\boldsymbol{\theta}}(kT - T) + \boldsymbol{P}_0(kT)\boldsymbol{\varphi}(kT)\big[y(kT) - \boldsymbol{\varphi}^{\mathrm{T}}(kT)\hat{\boldsymbol{\theta}}(kT - T)\big], \tag{7}$$

$$\boldsymbol{P}_0^{-1}(kT) = \boldsymbol{P}_0^{-1}(kT - T) + \boldsymbol{\varphi}(kT)\boldsymbol{\varphi}^{\mathrm{T}}(kT), \quad \boldsymbol{P}_0(0) = p_0\boldsymbol{I}, \tag{8}$$

where $\boldsymbol{P}_0(kT) \in \mathbb{R}^{n_0 \times n_0}$ is the covariance matrix and $\boldsymbol{I}$ is an identity matrix of appropriate sizes.

Remark 3 It is known that the RLS algorithm has the advantage of fast convergence rate, but its computational burden is relatively heavy. From (7)–(8), we can see that the RLS algorithm for non-uniformly sampled systems requires calculating the covariance matrix $\boldsymbol{P}_0(kT)$ of a large size $n_0 \times n_0$ at each recursion step. Since the number $n_0 = (r + 1)n$ is larger than the parameter number $2n + 1$ of single-rate systems, the RLS algorithm in (7)–(8) tolerates a heavier computational burden than the RLS algorithm for single-rate systems. This motivates us to present a highly computationally efficient algorithm to estimate the parameter vector $\boldsymbol{\theta}$ in (6).

4 The hierarchical estimation algorithm

The hierarchical identification principle is an effective way of handling large-scale systems by reducing the computational load [5–7, 11]. Here, we use the hierarchical identification principle to derive the hierarchical least squares (HLS) algorithm for the system in (5).

Decompose the information vector $\boldsymbol{\varphi}(kT)$ into N sub-information vectors and the parameter vector $\boldsymbol{\theta}$ into N sub-parameter vectors with dimension n_i, i.e.,

$$\boldsymbol{\varphi}(kT) = \begin{bmatrix} \boldsymbol{\varphi}_1(kT) \\ \boldsymbol{\varphi}_2(kT) \\ \vdots \\ \boldsymbol{\varphi}_N(kT) \end{bmatrix} \in \mathbb{R}^{n_0}, \quad \boldsymbol{\varphi}_i(kT) \in \mathbb{R}^{n_i},$$

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_N \end{bmatrix} \in \mathbb{R}^{n_0}, \quad \boldsymbol{\theta}_i \in \mathbb{R}^{n_i}, \ n_1 + n_2 + \cdots + n_N = n_0.$$

Then the overall identification model in (6) can be written as

$$y(kT) = \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{\theta}_i + \sum_{j=1, j\neq i}^{N} \boldsymbol{\varphi}_j^{\mathrm{T}}(kT)\boldsymbol{\theta}_j + v(kT). \tag{9}$$

Let $\hat{\boldsymbol{\theta}}_i(kT)$ be the estimate of $\boldsymbol{\theta}_i$ at time $t = kT$. According to the least squares principle, the estimation algorithm for $\boldsymbol{\theta}_i$ in (9) can be expressed as

$$\hat{\boldsymbol{\theta}}_i(kT) = \hat{\boldsymbol{\theta}}_i(kT - T) + \boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)$$

$$\times \left[y(kT) - \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\hat{\boldsymbol{\theta}}_i(kT - T) - \sum_{j\neq i}^{N} \boldsymbol{\varphi}_j^{\mathrm{T}}(kT)\boldsymbol{\theta}_j \right], \tag{10}$$

$$\boldsymbol{P}_i^{-1}(kT) = \boldsymbol{P}_i^{-1}(kT - T) + \boldsymbol{\varphi}_i(kT)\boldsymbol{\varphi}_i^{\mathrm{T}}(kT), \tag{11}$$

where $\boldsymbol{P}_i(kT)$ is the covariance matrix of the ith subsystem.

Remark 4 Because the expression on the right-hand side of (10) contains the unknown parameter vectors $\boldsymbol{\theta}_j$, $j = 1, 2, \ldots, i - 1, i + 1, \ldots, N$, using the hierarchical identification principle and replacing these unknown vectors $\boldsymbol{\theta}_j$ ($j \neq i$) in (10) with their estimates $\hat{\boldsymbol{\theta}}_j(kT - T)$ at the preceding time $t = kT - T$ yields

$$\hat{\boldsymbol{\theta}}_i(kT) = \hat{\boldsymbol{\theta}}_i(kT - T) + \boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)$$

$$\times \left[y(kT) - \sum_{j\neq i}^{N} \boldsymbol{\varphi}_j^{\mathrm{T}}(kT)\hat{\boldsymbol{\theta}}_j(kT - T) - \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\hat{\boldsymbol{\theta}}_i(kT - T) \right]$$

$$= \hat{\boldsymbol{\theta}}_i(kT - T) + \boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\left[y(kT) - \boldsymbol{\varphi}^{\mathrm{T}}(kT)\hat{\boldsymbol{\theta}}(kT - T) \right]. \tag{12}$$

Equations (12) and (11) form the decomposition-based least squares (HLS) identification algorithm.

To initialize the algorithm, we take $\boldsymbol{P}_i(0) = p_0 \boldsymbol{I}_{n_i}$, with p_0 being normally a large positive number (e.g., $p_0 = 10^6$) and $\hat{\boldsymbol{\theta}}_i(0) = \boldsymbol{1}_{n_i}/p_0$, with $\boldsymbol{1}_{n_i}$ being an n_i-dimensional column vector whose elements are all 1.

Remark 5 The diagram of the HLS algorithm is shown in Fig. 1. From the HLS algorithm in (11)–(12) and Fig. 1, we can see that the parameter estimate $\hat{\boldsymbol{\theta}}_i(kT)$ at time $t = kT$ depends not only on $\hat{\boldsymbol{\theta}}_i(kT - T)$ at the preceding time $t = kT - T$, but also on the estimates $\hat{\boldsymbol{\theta}}_j(kT - T)$ ($j = 1, 2, \ldots, i - 1, i + 1, \ldots, N$) of all the other submodels at time $t = kT - T$.

Remark 6 The proposed HLS algorithm in (11)–(12) has less computational burden than the RLS algorithm in (7)–(8). The computational burden of the two algorithms are compared in Table 1, where the numbers of multiplications and additions are for

Fig. 1 The diagram of the HLS algorithm
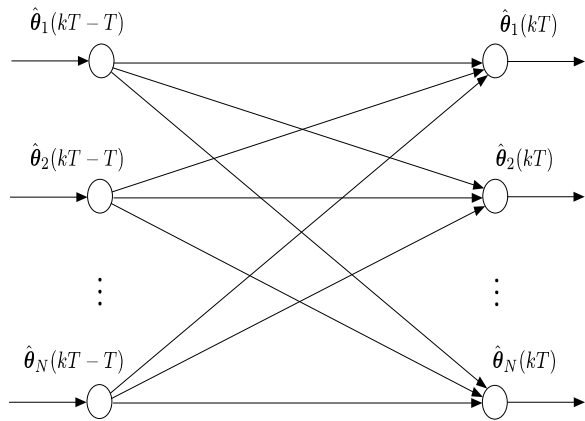
Table 1 Comparisons of computational efficiency

Algorithms	Number of multiplications	Number of additions
RLS	$2n_0^2 + 4n_0$	$2n_0^2 + 2n_0$
	{96}	{84}
HLS	$\sum_{i=1}^{N}[2n_i^2 + 4n_i]$	$\sum_{i=1}^{N}[2n_i^2 + 2n_i]$
	{60}	{48}

each step. Taking a non-uniformly sampled system with $n_0 = 6$ and $N = 2$ as an example, i.e., $n_1 = n_2 = 3$, the times of multiplication and addition are shown in the braces.

Remark 7 From Table 1, we can see that the computation cost of the HLS algorithm in (11)–(12) is lower than that of the RLS algorithm, and the computational cost of the HLS algorithm depends on the choice of the number N. The larger N, the lower the computation cost. When $N = n_0$, i.e., $n_i = 1$, we get the least computational cost.

5 Main convergence results

In this section, we establish the main convergence results of the HLS algorithm for non-uniformly sampled systems, using the martingale convergence theorem (Lemma D.5.3 in [16]). Two problems are addressed: (1) Do the parameter estimates given by the HLS algorithm converge to the true parameters? (2) How fast does the algorithm converge?

Let us introduce some notations. The symbols $\lambda_{\max}[X]$ and $\lambda_{\min}[X]$ represent the maximum and minimum eigenvalues of the positive definite matrix X, respectively; $\|X\|^2 := \mathrm{tr}[XX^T]$. The relation $f(k) = O(g(k))$ means that there exist positive constants δ_1 and k_0 such that $|f(k)| \leqslant \delta_1 g(k)$ for $g(k) \geqslant 0$ and $k \geqslant k_0$.

Assume that $\{v(kT), \mathcal{F}_{kT}\}$ is a martingale difference sequence defined on a probability space $\{\Omega, \mathcal{F}, P\}$, where $\{\mathcal{F}_{kT}\}$ is the σ algebra sequence generated by the observations up to and including time $t = kT$ [16]. The noise sequence $\{v(kT)\}$ satisfies the following assumptions:

$$(A1) \quad E[v(kT) \mid \mathcal{F}_{kT-T}] = 0, \quad \text{a.s.};$$

$$(A2) \quad E[v^2(kT) \mid \mathcal{F}_{kT-T}] = \sigma^2 < \infty, \quad \text{a.s.}$$

Lemma 1 *For the algorithm in* (11)–(12), *the following inequalities hold*:

$$\sum_{j=1}^{\infty} \frac{\varphi_i^{\mathrm{T}}(jT) P_i(jT) \varphi_i(jT)}{[\ln |P_i^{-1}(jT)|]^c} < \infty, \quad \text{a.s., } c > 1.$$

The proof can be done in a similar way to that of Lemma 1 in [24].

Theorem 1 *For the system in* (6) *and the HLS algorithm in* (11)–(12), *if*

$$\gamma(kT) := \left(1 - \sum_{i=1}^{N} \varphi_i^{\mathrm{T}}(kT) P_i(kT) \varphi_i(kT)\right) \tilde{y}^2(kT) + 2\sum_{i \neq j}^{N} \tilde{y}_i(kT) \tilde{y}_j(kT) \leqslant 0,$$

then we take $\hat{\boldsymbol{\theta}}(kT) = \hat{\boldsymbol{\theta}}(kT - T)$. *Define*

$$R(kT) := \sum_{i=1}^{N} \left[\ln \left|P_i^{-1}(kT)\right|\right]^c$$

and assume that (A1) *and* (A2) *hold, then for any* $c > 1$, *we have*

$$\left\|\hat{\boldsymbol{\theta}}_i(kT) - \boldsymbol{\theta}_i\right\|^2 = O\left(\frac{R(kT)}{\lambda_{\min}[P_i^{-1}(kT)]}\right), \quad i = 1, 2, \ldots, N, \text{ a.s.}$$

Proof Define the parameter estimation error vector $\tilde{\boldsymbol{\theta}}(kT)$ and a nonnegative definite functions $Q_i(kT)$ as

$$\tilde{\boldsymbol{\theta}}_i(kT) := \hat{\boldsymbol{\theta}}_i(kT) - \boldsymbol{\theta}_i, \tag{13}$$

$$Q_i(kT) := \tilde{\boldsymbol{\theta}}_i^{\mathrm{T}}(kT) P_i^{-1}(kT) \tilde{\boldsymbol{\theta}}_i(kT), \quad i = 1, 2, \ldots, N, \tag{14}$$

and

$$\tilde{y}_i(kT) := \varphi_i^{\mathrm{T}}(kT) \tilde{\boldsymbol{\theta}}_i(kT - T),$$

$$\tilde{y}(kT) := \varphi^{\mathrm{T}}(kT) \tilde{\boldsymbol{\theta}}(kT - T) = \sum_{i=1}^{N} \tilde{y}_i(kT).$$

Define the innovation

$$e(kT) := y(kT) - \varphi^{\mathrm{T}}(kT) \hat{\boldsymbol{\theta}}(kT - T). \tag{15}$$

Using (9), we have

$$e(kT) = -\boldsymbol{\varphi}^{\mathrm{T}}(kT)\tilde{\boldsymbol{\theta}}(kT-T) + v(kT) = -\tilde{y}(kT) + v(kT). \tag{16}$$

Inserting (12) into (14) gives

$$\begin{aligned}
Q_i(kT) &= \big[\tilde{\boldsymbol{\theta}}_i(kT-T) + \boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)e(kT)\big]^{\mathrm{T}}\boldsymbol{P}_i^{-1}(kT) \\
&\quad \times \big[\tilde{\boldsymbol{\theta}}_i(kT-T) + \boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)e(kT)\big] \\
&= \tilde{\boldsymbol{\theta}}_i^{\mathrm{T}}(kT-T)\boldsymbol{P}_i(kT)\tilde{\boldsymbol{\theta}}_i(kT-T) + 2\tilde{\boldsymbol{\theta}}_i^{\mathrm{T}}(kT-T)\boldsymbol{\varphi}_i(kT)e(kT) \\
&\quad + \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)e^2(kT) \\
&= \tilde{\boldsymbol{\theta}}_i^{\mathrm{T}}(kT-T)\big[\boldsymbol{P}_i^{-1}(kT-T) + \boldsymbol{\varphi}_i(kT)\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\big]\tilde{\boldsymbol{\theta}}_i(kT-T) \\
&\quad + 2\tilde{\boldsymbol{\theta}}_i^{\mathrm{T}}(kT-T)\boldsymbol{\varphi}_i(kT)e(kT) \\
&\quad + \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)e^2(kT) \\
&= Q_i(kT-T) + \tilde{y}_i^2(kT) + 2\tilde{y}_i(kT)e(kT) \\
&\quad + \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)e^2(kT).
\end{aligned} \tag{17}$$

Define the stochastic Lyapunov function

$$Q(kT) := \sum_{i=1}^{N} Q_i(kT).$$

Using (17) and (16), we have

$$\begin{aligned}
Q(kT) &= \sum_{i=1}^{N} Q_i(kT-T) + \sum_{i=1}^{N} \tilde{y}_i^2(kT) + 2\sum_{i=1}^{N} \tilde{y}_i(kT)e(kT) \\
&\quad + \sum_{i=1}^{N} \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)e^2(kT) \\
&= Q(kT-T) + \sum_{i=1}^{N} \tilde{y}_i^2(kT) + 2\tilde{y}(kT)\big[-\tilde{y}(kT) + v(kT)\big] \\
&\quad + \sum_{i=1}^{N} \boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\big[-\tilde{y}(kT) + v(kT)\big]^2 \\
&= Q(kT-T) + \sum_{i=1}^{N} \tilde{y}_i^2(kT) - 2\sum_{i=1}^{N} \tilde{y}_i^2(kT)
\end{aligned}$$

$$-4\sum_{i\neq j}^{N}\tilde{y}_i(kT)\tilde{y}_j(kT)+2\tilde{y}(kT)v(kT)$$

$$+\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\big\{\big[\tilde{y}(kT)\big]^2-2\tilde{y}(kT)v(kT)+v^2(kT)\big\}$$

$$=Q(kT-T)-\left[1-\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\right]\tilde{y}^2(kT)$$

$$-2\sum_{i\neq j}^{N}\tilde{y}_i(kT)\tilde{y}_j(kT)+2\left[1-\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\right]\tilde{y}(kT)v(kT)$$

$$+\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)v^2(kT)$$

$$\leqslant Q(kT-T)-\gamma(kT)$$

$$+2\left[1-\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\right]\sum_{i=1}^{N}\tilde{y}_i(kT)v(kT)$$

$$+\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)v^2(kT). \tag{18}$$

Since $\{v(kT)\}$ is a white noise sequence, taking the conditional expectation of both sides of (18) with respect to $\mathcal{F}_{kT-T}$ and using (A1)–(A2) give

$$\mathrm{E}\big[Q(kT)\mid\mathcal{F}_{kT-T}\big]=Q(kT-T)-\gamma(kT)+\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\sigma^2.$$

If $\gamma(kT)\geqslant 0$, we have

$$\mathrm{E}\big[Q(kT)\mid\mathcal{F}_{kT-T}\big]\leqslant Q(kT-T)+\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\sigma^2; \tag{19}$$

otherwise, we have $\hat{\boldsymbol{\theta}}(kT):=\hat{\boldsymbol{\theta}}(kT-T)$ and $Q(kT):=Q(kT-T)$. Thus, we always have

$$\mathrm{E}\big[Q(kT)\mid\mathcal{F}_{kT-T}\big]=Q(kT-T)+\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)\sigma^2. \tag{20}$$

Define

$$Z(kT):=\frac{Q(kT)}{R(kT)}.$$

Using (20), we have

$$Z(kT) \leqslant \frac{Q(kT-T)}{\sum_{i=1}^{N}[\ln|\boldsymbol{P}_i^{-1}(kT)|]^c} + \frac{\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)}{\sum_{i=1}^{N}[\ln|\boldsymbol{P}_i^{-1}(kT)|]^c}\sigma^2$$

$$\leqslant Z(kT-T) + \sum_{i=1}^{N}\frac{\boldsymbol{\varphi}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i(kT)\boldsymbol{\varphi}_i(kT)}{[\ln|\boldsymbol{P}_i^{-1}(kT)|]^c}\sigma^2. \tag{21}$$

According to Lemma 1, the summation of the last term of the right-hand side of (21) for k from 1 to ∞ is finite. Applying the martingale convergence theorem (Lemma D.5.3 in [16]) to (21), we conclude that $Z(kT)$ converges a.s. to a finite random variable, say Z_0, i.e.,

$$Z(kT) = \frac{Q(kT)}{R(kT)} \to Z_0 < \infty.$$

Thus we have

$$Q(kT) = O\big(R(kT)\big), \quad Q_i(kT) = O\big(R(kT)\big).$$

From the definition of $Q_i(kT)$, we have

$$\big\|\tilde{\boldsymbol{\theta}}_i(kT)\big\|^2 \leqslant \frac{\tilde{\boldsymbol{\theta}}_i^{\mathrm{T}}(kT)\boldsymbol{P}_i^{-1}(kT)\tilde{\boldsymbol{\theta}}_i(kT)}{\lambda_{\min}[\boldsymbol{P}_i(kT)]}$$

$$= \frac{Q_i(kT)}{\lambda_{\min}[\boldsymbol{P}_i(kT)]} = O\bigg(\frac{R(kT)}{\lambda_{\min}[\boldsymbol{P}_i(kT)]}\bigg), \quad i=1,2,\dots,N, \text{ a.s.}$$

This proves Theorem 1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 1 *For the system in* (6) *and the HLS algorithm in* (11)–(12), *assume that the conditions of Theorem* 1 *hold and that there exist positive constants* $c_1, c_2, c_0,$ *and* k_0 *such that for* $k \geqslant k_0,$ *the following general persistent excitation condition holds*:

$$\text{(A3)} \quad c_1\boldsymbol{I}_{n_i} \leqslant \frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\varphi}_i(jT)\boldsymbol{\varphi}_i^{\mathrm{T}}(jT) \leqslant c_2 k^{c_0}\boldsymbol{I}_{n_i}, \quad i=1,2,\dots,N, \text{ a.s.}$$

Then the parameter estimation errors converge to zero like $\|\tilde{\boldsymbol{\theta}}_i(kT)\|^2 \to 0$, *a.s.,* $i=1,2,\dots,N.$

Proof From the definitions of $\boldsymbol{P}_i(kT)$ and the condition (A3), we have

$$c_1 k\boldsymbol{I}_{n_i} \leqslant \boldsymbol{P}_i^{-1}(kT) \leqslant c_2 k^{c_0+1}\boldsymbol{I}_{n_i} + \frac{1}{p_0}\boldsymbol{I}_{n_i},$$

$$\lambda_{\min}\big[\boldsymbol{P}_i^{-1}(kT)\big] \geqslant c_1 k,$$

$$\ln\big|\boldsymbol{P}_i^{-1}(kT)\big| \leqslant n_i \ln\bigg[c_2 k^{c_0+1} + \frac{1}{p_0}\bigg].$$

Using Theorem 1, we have

$$\|\hat{\boldsymbol{\theta}}_i(kT) - \boldsymbol{\theta}_i\|^2 = O\left(\frac{R(kT)}{\lambda_{\min}[\boldsymbol{P}_i(kT)]}\right)$$

$$= O\left(\frac{1}{c_1 k}\sum_{i=1}^{N} n_i^c \left[\ln\left(c_2 k^{c_0+1} + 1/p_0\right)\right]^c\right)$$

$$= O\left(\frac{N[\ln k]^c}{k}\right) \to 0, \quad i = 1, 2, \ldots, N, \text{ a.s.}$$

The proof is completed.   □

Remark 8 Corollary 1 shows that the estimation errors $\|\tilde{\boldsymbol{\theta}}_i(kT)\|^2$ converge to zero at the rate of $N[\ln k]^c/k$ as k approaches infinity. This implies that the convergence rate depends on the number N and that the larger N, the slower the convergence rate. Therefore, the HLS algorithm in (11)–(12) reduces the computational load at the cost of slowing down the convergence rate. In order to guarantee a certain convergence rate, the number N should not be too large.

6 Example

Consider a stable continuous-time process with the following state-space representation:

$$S_c: \quad \begin{cases} \dot{\boldsymbol{x}}(t) = \begin{bmatrix} -0.1 & -0.5 \\ 1 & 0 \end{bmatrix}\boldsymbol{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u(t), \\ y(t) = [0, 2]\boldsymbol{x}(t). \end{cases}$$

Take $r = 2$, $\tau_1 = 0.3$ s, $\tau_2 = 0.5$ s, then $t_1 = \tau_1 = 0.3$ s, $t_2 = \tau_1 + \tau_2 = T = 0.8$ s. Discretizing this system gives

$$\begin{cases} x(kT+T) = \begin{bmatrix} 0.7754 & -0.3642 \\ 0.7285 & 0.8483 \end{bmatrix} x(kT) + \begin{bmatrix} 0.2509 & 0.4776 \\ 0.1818 & 0.1217 \end{bmatrix}\begin{bmatrix} u(kT) \\ u(kT+t_1) \end{bmatrix}, \\ y(kT) = [0, 2]x(kT). \end{cases}$$

The corresponding input–output relationship is

$$\left(1 - 1.6240z^{-1} + 0.9231z^{-2}\right)y(kT)$$

$$= \left(0.3636z^{-1} + 0.0836z^{-2}\right)u(kT)$$

$$+ \left(0.2433z^{-1} + 0.5072z^{-2}\right)u(kT+t_1) + v(kT).$$

In the simulation, the inputs $\{u(kT + t_i), i = 0, 1\}$ are taken as persistent excitation signal sequences with zero mean and unit variance, and $\{v(kT)\}$ as a white noise sequence with zero mean and variance $\sigma^2 = 0.10^2$. The corresponding noise-

Table 2 The RLS parameter estimates and errors

k	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	δ (%)
100	−1.61612	0.91343	0.35807	0.08953	0.23690	0.50639	0.81666
200	−1.61400	0.91200	0.36325	0.08797	0.23135	0.50325	1.00806
500	−1.61957	0.91958	0.36689	0.08817	0.23572	0.50105	0.63472
1000	−1.61987	0.91915	0.36622	0.08655	0.24118	0.49998	0.51537
2000	−1.62173	0.92164	0.36581	0.08308	0.24130	0.50172	0.34297
3000	−1.62271	0.92229	0.36476	0.08414	0.24004	0.50352	0.26680
True values	−1.62400	0.92310	0.36360	0.08360	0.24330	0.50720	

Table 3 The HLS parameter estimates and errors with $N = 2$

k	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	δ (%)
100	−1.60036	0.90200	0.19301	0.21073	0.12421	0.48222	12.44216
200	−1.59887	0.90083	0.27020	0.15677	0.18314	0.50219	6.91245
500	−1.61081	0.91322	0.33452	0.11324	0.21580	0.50743	2.64112
1000	−1.61335	0.91424	0.35027	0.10031	0.23041	0.50343	1.44977
2000	−1.61832	0.91892	0.35776	0.09146	0.23572	0.50457	0.72947
3000	−1.62060	0.92063	0.35939	0.09047	0.23616	0.50595	0.58537
True values	−1.62400	0.92310	0.36360	0.08360	0.24330	0.50720	
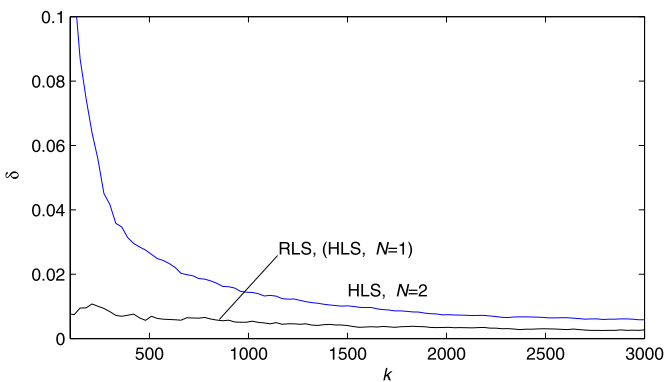
Fig. 2 The parameter estimation error δ versus k

to-signal ratio is $\delta_{ns} = 11.82$ %. The parameter estimates and their errors $\delta :=$ $\|\hat{\boldsymbol{\theta}}(kT) - \boldsymbol{\theta}\| / \|\boldsymbol{\theta}\|$ given by the RLS algorithm in (7)–(8) are shown in Table 2. In this simulation, the number of parameters is $n_0 = 6$. Let $N = 2$ and $n_1 = n_2 = 3$, then apply the HLS algorithm in (11)–(12) to estimate the parameters of this non-uniformly sampled system. The simulation results are shown in Table 3. The estimation errors δ versus k of the two algorithms are shown in Fig. 2.

From Tables 1–3 and Fig. 2, we have the following conclusions:

– The identification model is a fictitious two-input single-output system because the input is non-uniformly sampled twice over the frame period. Table 2 and Fig. 2 show that RLS algorithm can give highly accurate estimation.
– In this simulation, $r = 2$, $n_0 = 6$ and $N = 2$. From Table 1, we can see that the numbers of multiplication and addition at each step for the HLS algorithm are 60 and 48, respectively, and are fewer than 96 and 84 of the RLS algorithm. Compared to the RLS algorithm, the HLS algorithm saves 40 % computational cost.
– The parameter estimation errors for the HLS algorithm become smaller and gradually get close to zero as the data length k increases.
– The convergence rate of the HLS algorithm is a little slower than that of the RLS algorithm.

7 Conclusions

In this paper, an HLS algorithm is presented for a class of non-uniformly sampled systems based on the hierarchical identification principle. The main advantage of the proposed algorithm is that it can reduce the computation load compared with the traditional RLS algorithm. The convergence analysis indicates that the parameter estimates given by the HLS algorithm can converge to their true values, and the number N cannot be too large in the algorithm implementation because the number of the submodels really affects the convergence rate. A numerical example shows that the HLS algorithm can give effective parameter estimates with lower computational effort. The proposed hierarchical method based on decomposing the parameter vector and the information vector can be extended to a variety of system models such as the multivariable systems with a complicated colored noise [12]. Due to its benefit of reduced computational complexity, the proposed identification method could be applied to resource-constrained networked dynamic systems [19, 29, 30, 42, 44].

References

1. A. Al-Smadi, A least-squares-based algorithm for identification of non-Gaussian ARMA models. Circuits Syst. Signal Process. **26**(5), 715–731 (2007)
2. T. Chen, B. Francis, *Optimal Sampled-Data Control Systems* (Springer, London, 1995)
3. M. Cimino, P.R. Pagilla, Design of linear time-invariant controllers for multirate systems. Automatica **46**(8), 1315–1319 (2010)
4. M. Cimino, P.R. Pagilla, Conditions for the ripple-free response of multirate systems using linear time-invariant controllers. Syst. Control Lett. **59**(8), 510–516 (2010)
5. F. Ding, T. Chen, Hierarchical least squares identification methods for multivariable systems. IEEE Trans. Autom. Control **50**(3), 397–402 (2005)
6. F. Ding, T. Chen, Hierarchical identification of lifted state-space models for general dual-rate systems. IEEE Trans. Circuits Syst. I, Regul. Pap. **52**(6), 1179–1187 (2005)

7. F. Ding, J. Yang, Hierarchical identification of large scale systems. Acta Autom. Sin. **25**(5), 647–654 (1999)

8. F. Ding, L. Qiu, T. Chen, Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems. Automatica **45**(2), 324–332 (2009)

9. F. Ding, G. Liu, X.P. Liu, Partially coupled stochastic gradient identification methods for non-uniformly sampled systems. IEEE Trans. Autom. Control **55**(8), 1976–1981 (2010)

10. F. Ding, G. Liu, X.P. Liu, Parameter estimation with scarce measurements. Automatica **47**(8), 1646–1655 (2011)

11. J. Ding, F. Ding, X.P. Liu, G. Liu, Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data. IEEE Trans. Autom. Control **56**(11), 2677–2683 (2011)

12. F. Ding, Y.J. Liu, B. Bao, Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems. Proc. Inst. Mech. Eng., Part I, J. Syst. Control Eng. **226**(1), 43–55 (2012)

13. M. Embiruçu, C. Fontes, Multirate multivariable generalized predictive control and its application to a slurry reactor for ethylene polymerization. Chem. Eng. Sci. **61**(17), 5754–5767 (2006)

14. J. Feng, Z. Wang, M. Zeng, Recursive robust filtering with finite-step correlated process noises and missing measurements. Circuits Syst. Signal Process. **30**(6), 1355–1368 (2011)

15. H. Fujimoto, Y. Hori, High-performance servo systems based on multirate sampling control. Control Eng. Pract. **10**(7), 773–781 (2002)

16. G.C. Goodwin, K.S. Sin, *Adaptive Filtering Prediction and Control* (Prentice-Hall, Englewood Cliffs, 1984)

17. R.B. Gopaluni, A particle filter approach to identification of nonlinear processes under missing observations. Can. J. Chem. Eng. **86**(6), 1081–1092 (2008)

18. S.C. Kadu, M. Bhushan, R. Gudi, Optimal sensor network design for multirate systems. J. Process Control **18**(6), 594–609 (2008)

19. H. Li, Y. Shi, Robust H_∞ filtering for nonlinear stochastic systems with uncertainties and random delays modeled by Markov chains. Automatica **48**(1), 159–166 (2012)

20. D. Li, S.L. Shah, T. Chen, K.Z. Qi, Application of dual-rate modeling to CCR octane quality inferential control. IEEE Trans. Control Syst. Technol. **11**(1), 43–51 (2003)

21. W. Li, S.L. Shah, D. Xiao, Kalman filters in non-uniformly sampled multirate systems: for FDI and beyond. Automatica **44**(1), 199–208 (2008)

22. Y.J. Liu, F. Ding, Decomposition based least squares estimation algorithm for non-uniformly sampled multirate systems, in *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 Held Jointly with the 2009 28th Chinese Control Conference, CDC/CCC 2009*. Shanghai, China (2009)

23. X. Liu, J. Lu, Least squares based iterative identification for a class of multirate systems. Automatica **46**(3), 549–554 (2010)

24. Y.J. Liu, L. Xie, F. Ding, An auxiliary model based on a recursive least-squares parameter estimation algorithm for non-uniformly sampled multirate systems. Proc. Inst. Mech. Eng., Part I, J. Syst. Control Eng. **223**(4), 445–454 (2009)

25. Y.J. Liu, L. Yu, F. Ding, Multi-innovation extended stochastic gradient algorithm and its performance analysis. Circuits Syst. Signal Process. **29**(4), 649–667 (2010)

26. L. Ljung, *System Identification: Theory for the User* (Prentice-Hall, Englewood Cliffs, 1999)

27. B. Ni, D. Xiao, Identification of non-uniformly sampled multirate systems with application to process data compression. IET Control Theory Appl. **4**(6), 970–984 (2010)

28. J. Sheng, T. Chen, S.L. Shah, Generalized predictive control for non-uniformly sampled systems. J. Process Control **12**(8), 875–885 (2002)

29. Y. Shi, T. Chen, Optimal design of multichannel transmultiplexers with stopband energy and passband magnitude constraints. IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process. **50**(9), 659–662 (2004)

30. Y. Shi, B. Yu, Robust mixed H_2/H_∞ control of networked control systems with random time delays in both forward and backward communication links. Automatica **47**(4), 754–760 (2011)

31. M. Srinivasarao, S.C. Patwardhan, R.D. Gudi, Nonlinear predictive control of irregularly sampled multirate systems using blackbox observers. J. Process Control **17**(1), 17–35 (2007)

32. Y.S. Suh, Stability and stabilization of nonuniform sampling systems. Automatica **44**(12), 3222–3226 (2008)

33. A.K. Tanc, A.H. Kayran, Maximum entropy power spectrum estimation for 2-D multirate systems. Circuits Syst. Signal Process. **31**(1), 271–281 (2012)

34. D.N. Vizireanu, Generalizations of binary morphological shape decomposition. J. Electron. Imaging **16**(1), 1–6 (2007)

35. D.N. Vizireanu, Morphological shape decomposition interframe interpolation method. J. Electron. Imaging **17**(1), 1–5 (2008)
36. D.N. Vizireanu, S. Halunga, G. Marghescu, Morphological skeleton decomposition interframe interpolation method. J. Electron. Imaging **19**(2), 1–3 (2010)
37. L. Xie, H.Z. Yang, F. Ding, Modeling and identification for non-uniformly periodically sampled-data systems. IET Control Theory Appl. **4**(5), 784–794 (2010)
38. H. Yang, Y. Xia, P. Shi, Stabilization of networked control systems with nonuniform random sampling periods. Int. J. Robust Nonlinear Control **21**(5), 501–526 (2011)
39. B. Yu, Y. Shi, H. Huang, l_2–l_∞ Filtering for multirate systems based on lifted models. Circuits Syst. Signal Process. **27**(5), 699–711 (2008)
40. H. Zhang, Y. Shi, A. Saadat Mehr, Robust energy-to-peak filtering for networked systems with time-varying delays and randomly missing data. IET Control Theory Appl. **4**(12), 2921–2936 (2010)
41. H. Zhang, Y. Shi, A. Saadat Mehr, Robust weighted H_∞ filtering for networked systems with intermitted measurements of multiple sensors. Int. J. Adapt. Control Signal Process. **25**(4), 313–330 (2011)
42. H. Zhang, Y. Shi, A. Saadat Mehr, Robust static output feedback control and remote PID design for networked motor systems. IEEE Trans. Ind. Electron. **58**(12), 5396–5405 (2011)
43. H. Zhang, Y. Shi, A. Saadat Mehr, H. Huang, Robust energy-to-peak FIR equalization for time-varying communication channels with intermittent observations. Signal Process. **91**(7), 1651–1658 (2011)
44. H. Zhang, Y. Shi, A. Saadat Mehr, Robust H_∞ PID control for multivariable networked control systems with disturbance/noise attenuation. Int. J. Robust Nonlinear Control **22**(2), 183–204 (2012)
45. Y. Zhu, H. Telkamp, J. Wang, Q. Fu, System identification using slow and irregular output samples. J. Process Control **19**(1), 58–67 (2009)