Scanning the Issue ........................................................................................................................................ 1749

PAPERS

Sensor Selection for Structural Observability in Discrete Event Systems Modeled by Petri Nets ........................................... Y. Ru and C. N. Hadjicostis 1751
Architectures for Distributed Controller With Sub-Controller Communication Uncertainty .............................................................. V. Yadav, M. V. Salapaka, and P. G. Voulgaris 1765
Finite Controllability of Infinite-Dimensional Quantum Systems .............. A. M. Bloch, R. W. Brockett, and C. Rangan 1797
Quantum Dissipative Systems and Feedback Control Design by Interconnection ........ M. R. James and J. E. Gough 1806
Multiple Model Adaptive Control With Mixing ........................................ M. Kuipers and P. Ioannou 1822
Distributed Coverage Verification in Sensor Networks Without Location Information ............................................................ A. Tahbaz-Salehi and A. Jadbabaie 1837
Awareness Coverage Control Over Large-Scale Domains With Intermittent Communications ................................................. Y. Wang and I. I. Hussein 1850
Topological Identification in Networks of Dynamical Systems ................. D. Materassi and G. Innocenti 1860

TECHNICAL NOTES AND CORRESPONDENCE

Further Results on Global Controllability of Planar Nonlinear Systems .............. Y. Sun 1872
Synthesis of Over-Approximating Inference-Based Decentralized Supervisors for Discrete Event Systems ......................................................... S. Takai and R. Kumar 1881
Stabilizability of Controlled Lagrangian Systems of Two Degrees of Freedom and One Degree of Under-Actuation by the Energy-Shaping Method .............................................. D. E. Chang 1888

(Contents Continued on Back Cover)
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Space Design of Repetitive Controllers for Satisfying a Robust Performance Requirement</td>
<td>B. Demirel and L. Güvenç</td>
<td>1893</td>
</tr>
<tr>
<td>On the Characterization of Strict Positive Realness for General Matrix Transfer Functions</td>
<td>M. Corless and R. Shorten</td>
<td>1899</td>
</tr>
<tr>
<td>Efficient Model Predictive Control for Nonlinear Systems via Function Approximation Techniques</td>
<td>M. Canale, L. Fagiano, and M. Milanese</td>
<td>1911</td>
</tr>
<tr>
<td>Razumikhin-Type Theorems on $p$th Moment Exponential Stability of Impulsive Stochastic Delay Differential Equations</td>
<td>S. Peng and Y. Zhang</td>
<td>1917</td>
</tr>
<tr>
<td>Uniting Two Control Lyapunov Functions for Affine Systems</td>
<td>V. Andrieu and C. Prieur</td>
<td>1923</td>
</tr>
<tr>
<td>ITAE Optimal Sliding Modes for Third-Order Systems With Input Signal and State Constraints</td>
<td>A. Bartoszewicz and A. Nowacka-Leverton</td>
<td>1928</td>
</tr>
<tr>
<td>Linear Copositive Lyapunov Functions for Continuous-Time Positive Switched Systems</td>
<td>E. Fornasini and M. E. Valcher</td>
<td>1933</td>
</tr>
<tr>
<td>On Gaussian Optimal Smoothing of Non-Linear State Space Models</td>
<td>S. Särkkä and J. Hartikainen</td>
<td>1938</td>
</tr>
<tr>
<td>Further Results on Stabilization of Shock-Like Equilibria of the Viscous Burgers PDE</td>
<td>A. Smyshlyaev, T. Meurer, and M. Krstic</td>
<td>1942</td>
</tr>
<tr>
<td>On Almost Sure Stability of Hybrid Stochastic Systems With Mode-Dependent Interval Delays</td>
<td>L. Huang and X. Mao</td>
<td>1946</td>
</tr>
<tr>
<td>An Interactive Term Approach to Non-Parametric FIR Nonlinear System Identification</td>
<td>E.-W. Bai and M. Deistler</td>
<td>1952</td>
</tr>
<tr>
<td>Stability Analysis for a String of Coupled Stable Subsystems With Negative Imaginary Frequency Response</td>
<td>C. Cai and G. Hagen</td>
<td>1958</td>
</tr>
<tr>
<td>Decentralized Dynamic Output Feedback Control of Nonlinear Interconnected Systems</td>
<td>K. Kalsi, J. Lian, and S. H. Zak</td>
<td>1964</td>
</tr>
<tr>
<td>Partially Coupled Stochastic Gradient Identification Methods for Non-Uniformly Sampled Systems</td>
<td>F. Ding, G. Liu, and X. P. Liu</td>
<td>1976</td>
</tr>
</tbody>
</table>
Partially Coupled Stochastic Gradient Identification Methods for Non-Uniformly Sampled Systems

Feng Ding, Guangjun Liu, and Xiaoping Peter Liu

Abstract—This technical note addresses identification problems of non-uniformly sampled systems. For the input-output representation of non-uniform discrete-time systems, a partially coupled stochastic gradient (C-SG) algorithm is proposed to estimate the model parameters with high computational efficiency compared with the standard stochastic gradient (SG) algorithm. The analysis indicates that the partially C-SG algorithm can give more accurate parameter estimates than the SG algorithm. The parameter estimates obtained using the partially C-SG algorithm converge to their true values as the data length approaches infinity.

Index Terms—Discretization, multirate systems, non-uniform sampling, parameter estimation, recursive identification, stochastic gradient.

I. INTRODUCTION

It is often assumed for the traditional discrete-time sampled-data systems that the control input updating period equals the output sampling period, and such systems are called single-rate systems [1]. The multirate systems arise owing to various types of sensors, which could have different and long processing time (especially for chemical sensors), and each sensor may have its own sampling frequency. Studies of dual-rate/multirate systems have a long history and can be traced to Krane’s work in 1957 [2]. Earlier work on multirate systems includes the control of a binary distillation column [3], the state and parameter estimation of the antibiotic fermentation processes [4] and bioreactors [5], [6] and the adaptive output feedback control of a cart-crane system [7], etc.

In the area of process control, Li et al. applied dual-rate modelling to the octane quality inferential control of the catalytic reformer in a petroleum refinery [8]; Embirucu et al. presented a multirate multi-variable generalized predictive control algorithm and its application to a slurry reactor for ethylene polymerization [9]; Liu et al. discussed input-to-state stabilization for nonlinear dual-rate sampled-data systems via an approximate discrete-time model [10]; Yu et al. studied the framework period.

Another class of multirate sampling is termed as the non-uniform sampling, with the sampling intervals for each variable being non-equally spaced [11]. Non-uniform sampling has advantages over uniform sampling, such as always preserving controllability and observability in discretization, supposing that one describes a non-uniformly sampled system by a lifted state space model [12], [13].

This technical note studies the parameter identification problems for periodically non-uniformly sampled systems which contain the traditional discrete-time systems as special cases. In this literature, Ding et al. presented a hierarchical identification algorithm to estimate the parameters and states for such non-uniformly sampled systems [12]; Sheng et al. proposed a generalized predictive control scheme for such non-uniformly sampled systems [13]; Li et al. developed a Kalman filter for a non-uniformly sampled system and investigated a Kalman filter-based methodology for unified detection and isolation of sensor, actuator, and process faults in the non-uniformly sampled system [16]. For related work, see recent studies on identification of dual-rate systems by using the polynomial transformation technique [17], by using an auxiliary model identification method [18]–[20] and by using the lifting technique [21], [22].

The contributions of this technical note lie in that a partially coupled stochastic gradient (C-SG) algorithm is derived to estimate the parameters of the non-uniformly sampled systems based on the input-output representation in [12]. The proposed C-SG algorithm has a higher computational efficiency than the standard stochastic gradient (SG) algorithm. The analysis indicates that the parameter estimates given by the C-SG algorithm are more accurate than those by the standard SG algorithm.

The rest of the technical note is organized as follows. Section II gives the input-output representation of non-uniformly sampled systems. Section III discusses the gradient based estimation algorithms, including the standard stochastic gradient (SG) algorithm, the subsystem SG algorithm and the partially coupled SG algorithm. Section IV studies the convergence of the proposed C-SG algorithm. Section V provides an illustrative example to validate the proposed method. Finally, Section VI offers some concluding remarks.

II. THE REPRESENTATIONS FOR NON-UNIFORM SYSTEMS

Let us introduce some notations first.

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_n$</td>
<td>An $n$-dimensional column vector with entries 1.</td>
</tr>
<tr>
<td>$a ij [X]$</td>
<td>The adjoint matrix of $X$: $a ij [X] = (a ij [X])$,</td>
</tr>
<tr>
<td>$c(0)[X]$</td>
<td>The column vector formed by the columns of $X$.</td>
</tr>
<tr>
<td>$A = X$</td>
<td>$X$ is defined by $A$.</td>
</tr>
<tr>
<td>$A = X$</td>
<td>$A$ is defined by $X$.</td>
</tr>
<tr>
<td>$\hat{a} (kT)$</td>
<td>The estimate of $a$ at time $t = kT$.</td>
</tr>
<tr>
<td>$\hat{a} (kT)$</td>
<td>The estimate of $A$ at time $t = kT$.</td>
</tr>
<tr>
<td>$\hat{a} (kT)$</td>
<td>The estimate of $\theta$ at time $t = kT$.</td>
</tr>
<tr>
<td>$\lambda_{\text{min}} [X]$</td>
<td>The minimum eigenvalue of the square matrix $X$.</td>
</tr>
<tr>
<td>$\bigcirc$</td>
<td>The Kronecker product. If $A = [a_{ij}] \in \mathbb{R}^{m \times n}$, $B = [b_{ij}] \in \mathbb{R}^{p \times q}$, then $A \bigcirc B = [a_{ij} b_{ij}] \in \mathbb{R}^{(mp) \times (nq)}$.</td>
</tr>
</tbody>
</table>

Manuscript received July 29, 2009; revised March 31, 2010; accepted April 22, 2010. Date of publication May 18, 2010; date of current version July 30, 2010. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by the Canada Research Chair Program and in part by the National Natural Science Foundation of China (60973043). Recommended by Associate Editor X. Chen.

F. Ding is with the Control Science and Engineering Research Center and the Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, China (e-mail: fding@jiangnan.edu.cn; fding@sce.carleton.ca).

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

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).

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.2010.2050713
Referring to [12, Section 2], this non-uniformly sampled system is depicted in Fig. 1, where $P_i$ is a continuous-time process, $\mathcal{H}$ denotes a non-uniform zero-order hold with the following characteristics [12], [14]:

$$ u(t) = \begin{cases} u(kT), & kT \leq t < kT + t_1, \\ u(kT + t_1), & kT + t_1 \leq t < kT + t_2, \\ \vdots \\ u(kT + t_r), & kT + t_r \leq t < (k + 1)T, \end{cases}$$

where $t = kT + t_i$ are the updating and sampling instants and $T = \tau_1 + \tau_2 + \cdots + \tau_r = t_r$ is the frame period, $t_i = t_1 + \tau_1 = t_2 + \tau_2 + \cdots + \tau_r$, $t_0 = 0$; the control input $u$ is updated $r$ times at the instants $t = kT + t_i, (i = 0, 1, \ldots, r - 1)$ over each framework period $T$, and the output $y$ is sampled $r$ times at the instants $t = kT + t_i, (i = 0, 1, \ldots, r - 1)$ over each framework period $T$. $S$ denotes a non-uniform sampler with

$$ y(kT + t_i) = y(t), \quad i = 0, 1, \ldots, (r - 1), \quad k = 0, 1, 2, \ldots $$

Referring to [12, Section 2], this non-uniformly sampled system has the discrete-time state space model of the following form ($i = 1, 2, \ldots$):

$$ x(kT + T) = Gx(kT) + Fw(kT), \quad y(kT + t_i) = C_{i-1}x(kT) + D_{i-1}w(kT) $$

where $x(kT + t_i) \in \mathbb{R}^n$ is the state vector, $G \in \mathbb{R}^{n \times n}, F \in \mathbb{R}^{n \times r}, C_i \in \mathbb{R}^{1 \times n}$ and $D_i \in \mathbb{R}^{1 \times r}$ are constant matrices/vectors, and the non-uniformly stacked input vector

$$ w(kT) := \begin{bmatrix} u(kT) \\ u(kT + t_1) \\ \vdots \\ u(kT + t_{r-1}) \end{bmatrix} \in \mathbb{R}^r. $$

**Remark 1:** Equations (2) and (3) form the discrete-time state space model which establishes the mapping relation between the available input and output data $\{u(kT + t_i), y(kT + t_i)\}$ for the non-uniformly sampled system.

**Remark 2:** The model in (2) and (3) is a “fictitious $r$-inputs $r$-outputs system” (i.e., a special multivariable system [23], [24]), although its corresponding continuous-time system is a single-input single-output one. $w(kT) \in \mathbb{R}^r$ in (4) is a stacked input vector instead of real multiple inputs. Here, “inputs” and “outputs” are the sampled values of the same input $u(t)$ and output $y(t)$ at different instants $t = kT + t_i$, respectively.

Let $z$ be a forward shift operator ($z^{-1}$ be a backward shift operator) like $x(kT + t_i) = x(kT + T + t_i)$ and $z^{-1}x(kT + t_i) = x(kT - T + t_i)$. From (2) and (3), we have

$$ y(kT + t_i) = [C_{i-1} z^{-1}(I_n - G)]F + D_{i-1}w(kT) $$

$$ = [z^{-1}C_{i-1} + \alpha_i(I_n - G)]F + D_{i-1}w(kT) $$

$$ = \beta_0(z) - \alpha_i(z)w(kT), \quad i = 1, 2, \ldots, r $$

or

$$ \alpha(z)y(kT + t_{i-1}) = \beta_i(z)w(kT), \quad i = 1, 2, \ldots, r $$

where $\alpha(z)$ is the characteristic polynomial of order $n$ and $\beta_i(z)$ is a row vector polynomial with

$$ \alpha(z) := z^n - \alpha_1 z^{n-1} - \alpha_2 z^{n-2} - \cdots - \alpha_n, \quad \alpha_i \in \mathbb{R}^r, $$

$$ \beta_i(z) := z^n C_{i-1} + \alpha_i z^{n-1} - \alpha_2 z^{n-2} - \cdots - \alpha_n, \quad \beta_i \in \mathbb{R}^r. $$

The objective of this technical note is to present new coupled identification methods to estimate the parameters of the input-output representation in (6) for the non-uniformly sampled systems, based on the given non-uniform input-output data $\{u(kT + t_i), y(kT + t_i) : i = 0, 1, \ldots, r-1, k = 0, 1, 2, \ldots \}$.

**III. THE GRADIENT BASED ESTIMATION ALGORITHMS**

This section derives a new coupled stochastic gradient identification algorithm for the input-output representation in (6) on the basis of the well-known stochastic gradient algorithm.

Let $n_0 := (n + 1)r$. Define the stacked output vector

$$ y(kT) := \begin{bmatrix} y(kT) \\ y(kT + t_1) \\ \vdots \\ y(kT + t_{r-1}) \end{bmatrix} \in \mathbb{R}^{n_0}, \quad \psi_i(kT) \in \mathbb{R}^{n_0} $$

and the output information matrix $\psi(kT)$ and input information vector $\phi(kT)$ as

$$ \psi(kT) := [\psi_1(kT - T), \psi_2(kT - 2T), \ldots, \psi_r(kT - nT)] \in \mathbb{R}^{n_0} $$

$$ \phi(kT) := [\psi_1(kT)^T, \psi_2(kT)^T, \ldots, \psi_r(kT)^T]^T \in \mathbb{R}^{n_0} $$

Let $\psi_i(kT) \in \mathbb{R}^{n_0}$ be the $i$th row vector of $\psi(kT)$. Define the parameter vector $\alpha \in \mathbb{R}^n$ and parameter matrix $\theta \in \mathbb{R}^{n \times n}$ as

$$ \alpha := [\alpha_1, \alpha_2, \ldots, \alpha_n]^T \in \mathbb{R}^n, $$

$$ \theta := [\theta_1, \theta_2, \ldots, \theta_n] \in \mathbb{R}^{n \times n}, $$

$$ \psi_i := [\psi_{i1}, \psi_{i2}, \ldots, \psi_{in}] \in \mathbb{R}^n $$

Taking into account the disturbance noise $v_i(kT) \in \mathbb{R}^r$, (6) can be written as

$$ y_i(kT) + \psi_i(kT) \alpha = \theta_i^T \phi(kT) + v_i(kT), \quad i = 1, 2, \ldots, r $$

**Remark 3:** The identification model in (9) has the following features: 1) each subsystem contains a common parameter vector $\alpha$ and a different parameter vector $\theta_i$; and 2) each subsystem contains a common input information vector $\phi(kT)$ and a different output information vector $\psi_i(kT)$. This motivates us to study on new coupled stochastic gradient algorithm.

**A. The Stochastic Gradient Algorithm**

Let $v(kT) := [v_1(kT), v_2(kT), \ldots, v_r(kT)]^T \in \mathbb{R}^r$. Combining $r$ subsystems in (9) gives a matrix form [23], [24]

$$ y(kT) + \psi(kT) \alpha = \theta^T \phi(kT) + v(kT). $$

This identification model in (10) contains a parameter vector \( \alpha \in \mathbb{R}^n \) and a parameter matrix \( \theta \in \mathbb{R}^{n \times r} \). In order to identify \( \alpha \in \mathbb{R}^n \) and \( \theta \), the model in (10) needs to be transformed into a new form. Let

\[
\bar{\theta} := \begin{bmatrix} \alpha \\ \text{col}[\theta] \end{bmatrix} \in \mathbb{R}^{n+nr},
\]

\[
\bar{\Phi}(kT) := [-\psi(kT), I_r \otimes \varphi^T(kT)] \in \mathbb{R}^{n \times (n+nr)}.
\] (11)

Then we have

\[
\psi(kT) = \Phi(kT)\bar{\theta} + \nu(kT)
\] (12)

where \( \bar{\theta} \) in (12) contains all parameters of the whole system.

Define the criterion function

\[
J_1(\bar{\theta}) := \| \psi(kT) - \Phi(kT)\bar{\theta} \|^2.
\]

Using the negative gradient search, referring to [25], [26] and minimizing \( J_1 \) gives the stochastic gradient (SG) algorithm of estimating \( \bar{\theta} \)

\[
\tilde{\bar{\theta}}(kT) = \bar{\theta}(kT - T) + \frac{\Phi^T(kT)}{R(kT)} [\psi(kT) - \Phi(kT)\tilde{\bar{\theta}}(kT - T)]
\]

\[
R(kT) = R(kT - T) + \| \Phi(kT) \|^2, \quad R(0) = 1.
\] (13) (14)

Remark 4: The SG algorithm in (13)–(14) has a large computational load, which is to be reduced with a proposed coupled stochastic gradient algorithm to be presented, together with the comparisons in terms of computational efficiency.

The following gives the subsystem SG algorithm in order to derive the new coupled SG algorithms with computationally high efficiency.

B. The Stochastic Gradient Algorithm for Subsystems

Let

\[
\phi_i(kT) := \begin{bmatrix} -\psi_i^T(kT) \\ \varphi_i(kT) \end{bmatrix} \in \mathbb{R}^{n_i+nr_i}.
\] (15)

Equation (9) can be written as

\[
\psi_i(kT) = \phi_i^T(kT) \begin{bmatrix} \alpha_i \\ \theta_i \end{bmatrix} + \nu_i(kT).
\] (16)

Minimizing the criterion function

\[
J_2(\alpha_i, \theta_i) := \| \psi_i(kT) - \phi_i(kT) \begin{bmatrix} \alpha_i \\ \theta_i \end{bmatrix} \|^2, \quad i = 1, 2, \ldots, r
\]

gives the 
subsystem stochastic gradient (S-SG) algorithm of estimating \( \alpha_i \) and \( \theta_i \)

\[
\begin{bmatrix} \hat{\alpha}_i(kT) \\ \hat{\theta}_i(kT) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_i(kT - T) \\ \hat{\theta}_i(kT - T) \end{bmatrix} + \frac{\phi_i(kT)}{R_i(kT)} \\
\times \left( \psi_i(kT) - \phi_i^T(kT) \begin{bmatrix} \hat{\alpha}_i(kT - T) \\ \hat{\theta}_i(kT - T) \end{bmatrix} \right),
\]

\[
R_i(kT) = R_i(kT - T) + \| \phi_i(kT) \|^2, \quad R_i(0) = 1, i = 1, 2, \ldots, r.
\] (17) (18)

The parameter estimates \( \hat{\alpha}_i(kT) \) and \( \hat{\theta}_i(kT) \) of each subsystem in (17)–(18) are independent, namely, the parameter estimates \( \hat{\alpha}_i \) and \( \hat{\theta}_i \) of Subsystem \( i \) do not depend on the \( \hat{\alpha}_j \) and \( \hat{\theta}_j \) of Subsystem \( j \) for \( i \neq j \). For the sake of clarity, we use \( \hat{\alpha}_i(kT) \) as \( \hat{\alpha}_i(kT) \) in Subsystem \( i \) and then the S-SG algorithm in (17)–(18) can be equivalently written as

\[
\begin{bmatrix} \hat{\alpha}_i(kT) \\ \hat{\theta}_i(kT) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_i(kT - T) \\ \hat{\theta}_i(kT - T) \end{bmatrix} + \frac{\phi_i(kT)}{R_i(kT)} \\
\times \left( \psi_i(kT) - \phi_i^T(kT) \begin{bmatrix} \hat{\alpha}_i(kT - T) \\ \hat{\theta}_i(kT - T) \end{bmatrix} \right),
\]

\[
R_i(kT) = R_i(kT - T) + \| \phi_i(kT) \|^2, \quad i = 1, 2, \ldots, r.
\] (19) (20)

This implies that there is no coupling between the parameter estimates of each subsystem.

Remark 5: \( \hat{\theta}_i(kT) \) in (19) is the estimate of the parameter vector \( \theta_i \) of the \( i \)-th subsystem and \( \hat{\alpha}_i(kT) \) in (19) is the estimate of the common parameter vector \( \alpha \) in all subsystems. Moreover, it is clear that the estimates of \( \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r \) are independent, and yet the parameter vector \( \alpha \) has \( r \) estimates: \( \hat{\alpha}_1(kT), \hat{\alpha}_2(kT), \ldots, \hat{\alpha}_r(kT) \) at time \( t = kT \) because \( \alpha \) is estimated once in each subsystem, and then this will result in a large amount of redundancy about the estimate of \( \alpha \), although one can take their average as the estimate of \( \alpha \)

\[
\hat{\alpha}(kT) = \frac{\hat{\alpha}_1(kT) + \hat{\alpha}_2(kT) + \cdots + \hat{\alpha}_r(kT)}{r} \in \mathbb{R}^n.
\] (21)

This motivates us to study new coupled stochastic gradient approach in order to avoid the redundant estimates of \( \alpha \).

C. The Partially Coupled Stochastic Gradient Algorithm

For recursive estimation algorithms, it is a reasonable assumption that the estimates generally approach their true values with the increasing of \( t = kT \). That is, one may think that the estimate \( \hat{\alpha}_{i-1}(kT) \) at time \( t = kT \) is closer to the true \( \alpha \) than \( \hat{\alpha}_i(kT) \) at \( t = kT - T \) [27], [28]. By means of the idea of the Jacobi and Gauss–Seidel iterations [29], replacing \( \hat{\alpha}_i(kT - T) \) on the right-hand side of (19) with \( \hat{\alpha}_{i-1}(kT) \) for \( i = 2, 3, \ldots, r \) and replacing \( \hat{\alpha}_1(kT - T) \) on the right-hand side of (19) with \( \hat{\alpha}_1(kT - T) \) for \( i = 1 \) give the following partially coupled stochastic gradient (C-SG) algorithm:

\[
\begin{bmatrix} \hat{\alpha}_i(kT) \\ \hat{\theta}_i(kT) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_i(kT - T) \\ \hat{\theta}_i(kT - T) \end{bmatrix} + \frac{\phi_i(kT)}{R_i(kT)} \\
\times \left( \psi_i(kT) - \phi_i^T(kT) \begin{bmatrix} \hat{\alpha}_i(kT - T) \\ \hat{\theta}_i(kT - T) \end{bmatrix} \right),
\]

\[
R_i(kT) = R_i(kT - T) + \| \phi_i(kT) \|^2, \quad i = 2, 3, \ldots, r.
\] (22) (23)

and

\[
\begin{bmatrix} \hat{\alpha}_1(kT) \\ \hat{\theta}_1(kT) \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_1(kT - T) \\ \hat{\theta}_1(kT - T) \end{bmatrix} + \frac{\phi_1(kT)}{R_1(kT)} \\
\times \left( \psi_1(kT) - \phi_1^T(kT) \begin{bmatrix} \hat{\alpha}_1(kT - T) \\ \hat{\theta}_1(kT - T) \end{bmatrix} \right),
\]

\[
R_1(kT) = R_1(kT - T) + \| \phi_1(kT) \|^2, \quad R_1(0) = 1.
\] (24) (25)

The schematic diagram of the partially coupled stochastic gradient algorithm in (22)–(25) is shown in Fig. 2 where \( \hat{\alpha}_i(kT) \) is the coupled parameter estimation vector, but \( \hat{\theta}_i(kT) \) is not. Thus this algorithm is called the partial coupled one. The estimates of \( \alpha, \theta_1, \ldots, \theta_r \) are \( \hat{\alpha}_i(kT) := \hat{\alpha}_i(kT), \hat{\theta}_1(kT), \ldots, \hat{\theta}_i(kT) \) at time \( t = kT \).

Remark 6: The C-SG algorithm in (22)–(25) has higher computational efficiency and avoids the redundant estimates of \( \alpha \), compared with the SG algorithm in (13)–(14) and the S-SG algorithm in
3) Compute information vectors

\[ \bar{\alpha}_i(kT - T) \]

by (24) and the estimates \( \hat{\alpha}_i(kT) \) by (22).

4) for \( i = 2 : r \) Compute \( R_i(kT) \) by (23) and the estimates \( \hat{\alpha}_i(kT) \) and \( \hat{\theta}_i(kT) \) by (22).

end

5) Increase \( k \) by 1 and go to Step 2.

The flowchart of computing the parameter estimates \( \hat{\alpha}_i(kT) \) and \( \hat{\theta}_i(kT) \) (\( i = 1, 2, \ldots, r \)) by the C-SG algorithm is shown in Fig. 3.

In order to improve the convergence rate of stochastic gradient algorithms, a forgetting factor \( 0 < \lambda < 1 \) is introduced into (14), (20), (23) and (25) as follows:

\[ R(kT) = \lambda R(kT - T) + ||\Phi(kT)||^2, \quad R(0) = 1, \quad (26) \]

\[ R_i(kT) = \lambda R_i(kT - T) + ||\phi_i(kT)||^2, \quad i = 1, 2, \ldots, r \quad (27) \]

we obtain the SG algorithm with \( \lambda \) in (13) and (26), the S-SG algorithm with \( \lambda \) in (19) and (27) and the C-SG algorithm with \( \lambda \) in (22), (24) and (27).

\begin{table}[h]
\centering
\caption{Comparison of Computational Efficiency.}
\begin{tabular}{|c|c|c|}
\hline
Algorithms & Number of multiplications & Number of additions \\
\hline
SG & \( 3r(n + n_0) + r \) & \( 3r(n + n_0) \) \\
S-SG & \( 3r(n + n_0) + 2r \) & \( 3r(n + n_0) + (r - 1)n \) \\
C-SG & \( 3r(n + n_0) + r \) & \( 3r(n + n_0) \) \\
\hline
\end{tabular}
\end{table}

IV. CONVERGENCE ANALYSIS

This section analyzes the convergence properties of the partially coupled stochastic gradient algorithm in (22)–(25). The convergence of the standard stochastic gradient algorithm in (13)–(14) has been reported in [25].

Assume that \( \{v_i(kT), \mathcal{F}_{kT}\} \ (i = 1, 2, \ldots, r) \) is a martingale difference sequence defined on a probability space \( \{\Omega, \mathcal{F}, P\} \), where \( \mathcal{F}_{kT} \) is the \( \sigma \) algebra sequence generated by \( \{v_i(kT)\} \).
Then the parameter estimates converge to their true values, i.e.,
\[ \lim_{T \to \infty} \hat{\theta}(kT) = \theta^* \]
and assume that (A1) and (A2) hold and
\[ \limsup_{k \to \infty} \frac{R_i(kT)}{\lambda_{\min}(Q(kT))} < \infty, \quad \text{a.s.} \]
Then the parameter estimates converge to their true values, i.e.,
\[ \hat{\theta}_i(kT) \to \alpha_i, \quad \text{a.s. and } \hat{\theta}_i(kT) \to \beta_i, \quad \text{a.s.} \]
Theorem 1 can be proved using a similar method given in [20] and is omitted here in order to save space.

V. EXAMPLE

Assume that the process model \( P_e \) has the following transfer function:
\[ P_e(s) = \frac{2s + 0.8}{s^2 + 0.8s + 0.8}. \]
This is a second-order system and its corresponding state space realization is given by
\[
\begin{align*}
\dot{x}(t) &= \begin{bmatrix} -0.8 & 0.8 \\ 1 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \\
y(t) &= \begin{bmatrix} 2 & 0.8 \end{bmatrix} x(t).
\end{align*}
\]
Let \( r = 2, t_0 = 0, t_1 = \sqrt{2} - 1, s_2 = T = 1 \), i.e., \( \tau_1 = t_1, s_2 = 2 - \sqrt{2}s \). Referring to the method in [12], discretizing this example system gives
\[
x(kT + T) = \begin{bmatrix} 0.2950 & -0.48066 \\ 0.60107 & 0.70714 \end{bmatrix} x(kT) + \begin{bmatrix} 0.15443 & 0.44665 \\ 0.22129 & 0.14440 \end{bmatrix} u(kT) + \begin{bmatrix} 2 & 1.60243 \\ 0.0 & 0.19999 \end{bmatrix} x(kT) + \begin{bmatrix} 0 \\ 1.75011 \end{bmatrix} u(kT + t_1).
\]
Assume that the corresponding non-uniform discrete-time system to be identified has the following form:
\[ \alpha(z) y(kT + t_1 - 1) = \beta(z) u(kT + t_1) + w(kT), \quad i = 1, 2, \]
\[ \alpha(z) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} \in \mathbb{R}^2, \]
\[ \beta(z) = \beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} \in \mathbb{R}^{1 \times 2}, \quad i = 1, 2. \]
The identification procedure is summarized as follows: First, we use the \( \text{idinput} \) function in Matlab to generate a random signal sequence with zero mean and unit variance as the input signals \( \{ u(kT), u(kT + t_1) \} \) and two uncorrelated noise sequences with zero mean and variances \( \sigma^2 = 0.10^2 \) as \( v_1(kT) \) and \( v_2(kT) \), and then compute the system outputs \( y(kT) \) and \( y(kT + t_1) \), the noise-to-signal ratio is \( \delta = \alpha_{0.10} = 11.59\% \). Second, based on the input and noisy output data, applying the SG, S-SG and C-SG algorithms with \( \lambda = 0.90 \) to estimate the parameters of this discrete-time model, the parameter estimates and their errors with different data lengths \( k \) are shown in Tables II–IV and the parameter estimation errors \( \delta \) versus \( k \) are shown in Fig. 4, where \( \delta : = \| \hat{\theta}(kT) - \theta \| / \| \theta \|, \quad \hat{\theta}(kT) \) denotes the estimate of the true parameter vector \( \theta \).

From Tables II–IV and Fig. 4, we can see that the estimation errors given by the SG, S-SG and C-SG algorithm become smaller with the data length \( k \) increasing, and among the three algorithms, the C-SG algorithm is the most accurate one and requires the least computational load because the step-sizes \( 1/R_i(kT) \) in the C-SG and S-SG algorithms are larger than the step-size \( 1/R(kT) \) in the SG algorithm.

In this example, the real-part of poles is \(-0.4\), and the sampling frequency of \( 1/Hz \) is quite slow. The imaginary part is \( 0.8 \), so the sampling frequency is relatively low, which could explain the slow convergence, as with \( k = 3000 \). There are 6000 measurements used for estimating the parameters \( \alpha \) and \( \beta \). Therefore, if we increase the sampling frequency, then the convergence rate can become faster.

The use of a forgetting factor in the gradient algorithms can improve the convergence rates. A smaller forgetting factor will increase the variance of the parameter estimates, and a larger forgetting factor leads to larger estimation errors. In other words, if we decrease the forgetting factor, the convergence rate of the parameter estimation will increase initially, but the variance of the estimation errors become larger with the increase of \( k \). Therefore, a solution to this problem is to adopt a time-varying forgetting factor; use a smaller forgetting factor at the initial period of the operation, and then let the forgetting factor gradually increase with \( k \), and finally approach 1 so that more accurate parameter estimates can be obtained. Reference [24] explained the influences...
of using a time-varying forgetting factor in the hierarchical stochastic gradient algorithm in the example section.

VI. CONCLUSION

This technical note presents a partially coupled stochastic gradient (C-SG) algorithm for non-uniformly sampled systems. The performance analysis indicates that the parameter estimates obtained using the proposed algorithms converge to their true values. The proposed C-SG algorithm can overcome the abundant estimates of the characteristic polynomial of the system and requires smaller computational load compared with the standard stochastic gradient algorithm. The analytical results have been confirmed with the simulation example.

REFERENCES