Improved Redundant Rule-Based Stochastic Gradient Algorithm for Time-Delayed Models Using Lasso Regression

HANGTAO ZHAO, LIXIN LV, AND YUEJIANG JI
Wuxi Vocational College of Science and Technology, Wuxi 214122, China

Corresponding author: Yuejiang Ji (jyj1981917@126.com)

ABSTRACT This paper proposes an improved redundant rule based lasso regression stochastic gradient (RR-LR-SG) algorithm for time-delayed models. The improved SG algorithm can update the parameter elements with different step-sizes and directions, thus it is more adaptive; while the lasso regression method can pick out the small weights from the redundant parameter vector, it therefore can obtain the time-delay easily. To show the effectiveness of the proposed algorithm, the convergence analysis is also given. The simulated numerical results are consistent with the analytically derived results of the proposed algorithm.

INDEX TERMS Parameter estimation, time-delayed model, lasso regression, redundant rule, stochastic gradient algorithm.

I. INTRODUCTION

In this paper, we consider the following time-delayed model,

$$y(t) = \boldsymbol{\alpha}(z)y(t - \tau) + \boldsymbol{\beta}(z)u(t) + v(t), \qquad (1)$$

where the polynomials $\boldsymbol{\alpha}(z)$ and $\boldsymbol{\beta}(z)$ are expressed as

$$\boldsymbol{\alpha}(z) = \alpha_1 z^{-1} + \cdots + \alpha_n z^{-n},$$
$$\boldsymbol{\beta}(z) = \beta_1 z^{-1} + \cdots + \beta_m z^{-m},$$

and the time-delay τ is unknown. The focus of this paper is to use an improved stochastic gradient algorithm to identify the unknown parameters and time-delay simultaneously through lasso regression.

SG algorithm, a special kind of gradient descent (GD) algorithm, is usually regarded as a worthy addition to the least squares (LS) algorithm [1]–[6]. It constitutes of two steps: the direction designing and the step-size calculating [7], [8]. Although, it does not require to compute a matrix inverse and an analytic function, its convergence rates are quite slow. To improve the convergence rates, some improved SG algorithms are developed. These algorithms, e.g., the multi-innovation SG algorithm [9], [10], the forgetting factor SG algorithm and the momentum SG algorithm [11]–[14], improve the efficiency of the traditional SG algorithm in two

The associate editor coordinating the review of this manuscript and approving it for publication was Lorenzo Mucchi.

ways: choose a better direction and calculate an optimal step-size. Note that all the elements in the parameter vector have the same step-size when starting the T-SG algorithm, which is unreasonable because each element in the parameter vector has its own order of magnitude. Therefore, assigning different step-sizes for each element is a better choice.

Time-delayed systems widely exist in engineering practices [15]–[17]. The identification of time-delayed system is more challenging because an unknown time-delay τ leads to an unknown information vector, which makes the identification algorithms be impossible. To overcome this difficulty, the iterative algorithm is often used. The basic idea of the iterative algorithm is to obtain the time-delay estimates and the parameter estimates iteratively [18], [19]. For example, Zhao et al developed a variational Bayesian (VB) approach for ARX models with a Markov chain time-varying time-delays, while the time-delay estimates and the parameters are obtained in the VB-E step and VB-M step, respectively [20]. Chen et al proposed an expectation maximization identification algorithm for time-delayed two-dimensional systems, where the time-delays and the parameters are updated iteratively [21]. Since the parameter estimates are depended on the time-delay estimates, and vice versa. Once one kind estimates have poor estimation accuracy, the other one will also have poor performance or may be divergent.

H. Zhao *et al.*: Improved Redundant Rule-Based Stochastic Gradient Algorithm for Time-Delayed Models Using Lasso Regression

IEEE *Access*

In this paper, an improved SG algorithm is proposed for an time-delayed model using the lasso regression. First, the redundant rule method is introduced to transform the time-delayed model into an augmented model whose parameter vector contains two zero sub-vectors and a parameter sub-vector. Then, the improved SG algorithm combining the lasso regression can distinguish these two kinds sub-vectors and then get the parameter and time-delay estimates simultaneously. The improved SG algorithm updates the parameter elements adaptively with different step-sizes and directions, therefore, it is more effective than the traditional SG algorithm.

Briefly, the paper is organized as follows. Section II describes the time-delayed model and the traditional SG algorithm. Section III introduces the framework of the redundant rule based SG algorithm. Section IV studies the RR-LR-SG algorithm. Section V provides an illustrative example. Finally, concluding remarks are given in Section VI.

II. THE TIME-DELAYED MODEL AND TRADITIONAL SG ALGORITHM

Rewrite the time-delayed model as a regression model

$$y(t) = \boldsymbol{\phi}^{\mathrm{T}}(t - \tau)\boldsymbol{\vartheta} + v(t),$$
$$\boldsymbol{\phi}(t - \tau) = [y(t - \tau - 1), \cdots, y(t - \tau - n), u(t - 1), \cdots,$$
$$u(t - m)]^{\mathrm{T}} \in \mathbb{R}^{n+m},$$
$$\boldsymbol{\vartheta} = [\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m]^{\mathrm{T}} \in \mathbb{R}^{n+m}.$$

Define the cost function

$$J(\boldsymbol{\vartheta}) = \frac{1}{2}[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t - \tau)\boldsymbol{\vartheta}]^2.$$

The parameter estimates by using the traditional SG (T-SG) algorithm are written by

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t - 1) + \frac{1}{r(t)}[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t - \tau)\hat{\boldsymbol{\vartheta}}(t - 1)],$$

where $\frac{1}{r(t)}$ is the step-size and can be computed by

$$r(t) = r(t - 1) + \boldsymbol{\phi}^{\mathrm{T}}(t - \tau)\boldsymbol{\phi}(t - \tau).$$

However, the T-SG algorithm is invalid for its unknown information vector $\boldsymbol{\phi}(t - \tau)$. To get the parameter estimates, one should assume that the time-delay τ is known in prior. The existing methods often use the iterative algorithm to update the parameters and time-delay [22], [23]. For example, in the sampling instant t, first let the time-delay lay between an interval $[0, M]$. Substitute the time-delay $\tau = i$, $i = 0, 1, \cdots, M$ into the following function

$$[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t - i)\hat{\boldsymbol{\vartheta}}(t - 1)]^2.$$

Assume that

$$j = \arg\min_i\{[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t - i)\hat{\boldsymbol{\vartheta}}(t - 1)]^2, i = 0, 1, \cdots, M\}.$$

Then we can get that the time-delay $\tau = j$ is the most likely value. Next, using the following SG algorithm to update the parameters yields

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t - 1) + \frac{1}{r(t)}[y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t - j)\hat{\boldsymbol{\vartheta}}(t - 1)].$$

These two steps iteratively run until the parameters and time-delay converge to the true values. Since the two steps are coupled with each other, one kind poor estimates can also lead to their corresponding estimates poor or divergent. A question naturally arises: can we develop a method which can estimate the parameters and time-delay simultaneously?

III. THE FRAMEWORK OF THE REDUNDANT RULE BASED SG ALGORITHM

The redundant rule method is an effective algorithm which can deal with system with invariant time-delay [24], [25]. This method can get the parameter estimates and time-delay estimates simultaneously. In the redundant rule method, assume that the upper bound of the time-delay τ is M. Then, the time-delayed model can be transformed into an augmented model

$$y(t) = \bar{\boldsymbol{\phi}}^{\mathrm{T}}(t - M)\bar{\boldsymbol{\vartheta}} + v(t),$$

where

$$\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t) = [\overbrace{y(t - 1), \cdots, y(t - \tau)}^{redundant\ vector},$$
$$\underbrace{y(t - \tau - 1), \cdots, y(t - \tau - n)}_{original\ vector},$$
$$\overbrace{y(t - \tau - n - 1), \cdots, y(t - M - n - 1)}^{redundant\ vector},$$
$$u(t - 1), \cdots, u(t - m)]^{\mathrm{T}} \in \mathbb{R}^{M+n+m},$$
$$\bar{\boldsymbol{\vartheta}} = [\overbrace{\bar{\alpha}_1, \cdots, \bar{\alpha}_\tau}^{zero\ vector}, \underbrace{\bar{\alpha}_{\tau+1}, \cdots, \bar{\alpha}_{\tau+n}}_{parameter\ vector},$$
$$\overbrace{\bar{\alpha}_{\tau+n+1}, \cdots, \bar{\alpha}_{M+n}}^{zero\ vector}, \beta_1, \cdots, \beta_m]^{\mathrm{T}} \in \mathbb{R}^{M+n+m}.$$

Since the redundant vectors play no role on the output $y(t)$, their corresponding parameter vectors are equal to zero vectors.

Using the redundant rule based SG algorithm to update the parameter vector $\bar{\boldsymbol{\vartheta}}$ yields

$$\bar{\boldsymbol{\vartheta}}(t) = \bar{\boldsymbol{\vartheta}}(t - 1) + \frac{1}{\bar{r}(t)}[y(t) - \bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)\bar{\boldsymbol{\vartheta}}(t - 1)],$$
$$\bar{r}(t) = \bar{r}(t - 1) + \bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)\bar{\boldsymbol{\phi}}(t).$$

Let

$$\bar{\boldsymbol{\vartheta}}(t) = [\bar{\vartheta}_1(t), \cdots, \bar{\vartheta}_{M+n+m}(t)]^{\mathrm{T}} \in \mathbb{R}^{M+n+m}.$$
$$\bar{\boldsymbol{\phi}}(t) = [\bar{\phi}_1(t), \cdots, \bar{\phi}_{M+n+m}(t)]^{\mathrm{T}} \in \mathbb{R}^{M+n+m}.$$

When the estimates $\bar{\boldsymbol{\vartheta}}(t)$ have been obtained, we compare the elements $\bar{\vartheta}_i(t)$ with a threshold ζ. If $|\bar{\vartheta}_i(t)| < \zeta$, this element

can be regarded as a redundant element and be picked out from the vector $\bar{\boldsymbol{\vartheta}}(t)$. Once all the redundant elements have been picked out from the redundant parameter vector, the time-delay and the true parameter estimates can be obtained simultaneously.

Remark 1: The threshold ζ plays an important role in the redundant rule based SG algorithm. If the threshold ζ is small, some redundant elements will be not picked out from the parameter vector $\boldsymbol{\vartheta}$. On the other hand, a large threshold ζ would lead some true elements be mistaken for redundant ones. In application, we usually choose the threshold on a case by case basis.

IV. THE REDUNDANT RULE BASED LASSO REGRESSION SG ALGORITHM

The lasso regression method is usually applied to overcome overfitting, where some secondary factors are involved in the structure of the system. In this paper, the lasso regression method is extended to the time-delayed model to pick out the zero vectors from the parameter vector.

A. ALGORITHM DESIGNING

To use the lasso regression method for the time-delayed model, first let us introduce the following lemma.

Lemma 1: For the function $y = |x|$, its first derivative is

$$|x|' = \begin{cases} 1, & \text{if } x > 0, \\ [-1,1], & \text{if } x = 0, \\ -1, & \text{if } x < 0 \end{cases}.$$

Define the cost function as

$$J(\bar{\boldsymbol{\vartheta}}) = J_1(\bar{\boldsymbol{\vartheta}}) + J_2(\bar{\boldsymbol{\vartheta}}) = \frac{1}{2}[y(t) - \bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)\bar{\boldsymbol{\vartheta}}]^2 + \frac{\lambda}{2}\|\bar{\boldsymbol{\vartheta}}^{\mathrm{T}}\|_1,$$

where

$$J_1(\bar{\boldsymbol{\vartheta}}) = \frac{1}{2}[y(t) - \bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)\bar{\boldsymbol{\vartheta}}]^2$$
$$= \frac{1}{2}[y(t) - \bar{\phi}_1(t)\bar{\vartheta}_1 - \cdots - \bar{\phi}_{M+n+m}(t)\bar{\vartheta}_{M+n+m}]^2,$$
$$J_2(\bar{\boldsymbol{\vartheta}}) = \frac{\lambda}{2}|\bar{\vartheta}_1| + \cdots + \frac{\lambda}{2}|\bar{\vartheta}_{M+n+m}|.$$

By using the regularization term $\frac{\lambda}{2}\|\bar{\boldsymbol{\vartheta}}^{\mathrm{T}}\|_1$, some redundant parameters can be picked out from the augmented parameter vector, and then the time-delay can be estimated based on the structure of the new parameter vector.

Taking the derivative of $J_1(\bar{\boldsymbol{\vartheta}})$ with respect to $\bar{\vartheta}_1$ yields

$$\frac{J_1(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)} = -\bar{\phi}_1(t)[y(t) - \bar{\phi}_1(t)\bar{\vartheta}_1(t-1) - \cdots$$
$$- \bar{\phi}_{M+n+m}(t)\bar{\vartheta}_{M+n+m}(t-1)]. \quad (2)$$

Taking the derivative of $J_2(\bar{\boldsymbol{\vartheta}})$ with respect to $\bar{\vartheta}_1$ yields

$$\frac{J_2(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)} = \begin{cases} \frac{\lambda}{2}, & \text{if } \bar{\vartheta}_1 > 0, \\ [-a,a], & \text{if } \bar{\vartheta}_1 = 0, \\ -\frac{\lambda}{2}, & \text{if } \bar{\vartheta}_1 < 0 \end{cases}, \quad (3)$$

where $a = \frac{1}{2}\lambda$.

Then, the parameter element $\bar{\vartheta}_1$ in the sampling instant t can be computed by

$$\bar{\vartheta}_1(t) = \bar{\vartheta}_1(t-1) - r\left[\frac{J_1(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)} + \frac{J_2(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)}\right]$$
$$= \bar{\vartheta}_1(t-1) + r\bar{\phi}_1(t-1)[y(t) - \bar{\phi}_1(t)\bar{\vartheta}_1(t-1) - \cdots$$
$$- \bar{\phi}_{M+n+m}(t)\bar{\vartheta}_{M+n+m}(t-1)] - r\frac{J_2(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)}.$$

Substituting Equation (3) into the above equation yields

$$\bar{\vartheta}_1(t)$$
$$= \bar{\vartheta}_1(t-1) - r\left[\frac{J_1(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)} + \frac{J_2(\bar{\boldsymbol{\vartheta}})}{\bar{\vartheta}_1}\Big|_{\bar{\boldsymbol{\vartheta}}(t-1)}\right]$$
$$= \begin{cases} \bar{\vartheta}_1(t-1) + r\bar{\phi}_1(t)[y(t) - \bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\vartheta}}(t-1)] - r\frac{\lambda}{2}, \\ \quad \text{if } m_1 > \frac{\lambda}{2}, \\ 0, \text{ if } m_1 \in [-\frac{\lambda}{2}, \frac{\lambda}{2}] \\ \bar{\vartheta}_1(t-1) + r\bar{\phi}_1(t)[y(t) - \bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\vartheta}}(t-1)] + r\frac{\lambda}{2}, \\ \quad \text{if } m_1 < -\frac{\lambda}{2} \end{cases},$$

where

$$m_1(t) = \bar{\phi}_1(t)[y(t) - \bar{\phi}_2(t)\bar{\vartheta}_2(t-1) - \cdots$$
$$- \bar{\phi}_{M+n+m}(t)\bar{\vartheta}_{M+n+m}(t-1)]. \quad (4)$$

Therefore, we can get the redundant rule based lasso regression SG (RR-LR-SG) algorithm

$$\bar{\vartheta}_i(t) = \begin{cases} \bar{\vartheta}_i(t-1) + r_i\bar{\phi}_i(t)[y(t) - \bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\vartheta}}(t-1)] - r_i\frac{\lambda}{2}, \\ \quad \text{if } m_i(t) > \frac{\lambda}{2}, \\ 0, \text{ if } m_i(t) \in [-\frac{\lambda}{2}, \frac{\lambda}{2}] \\ \bar{\vartheta}_i(t-1) + r_i\bar{\phi}_i(t)[y(t) - \bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\vartheta}}(t-1)] + r_i\frac{\lambda}{2}, \\ \quad \text{if } m_i < -\frac{\lambda}{2} \end{cases},$$
$$m_i(t) = \bar{\phi}_i(t)[y(t) - \bar{\phi}_1(t)\bar{\vartheta}_1(t-1) - \cdots$$
$$- \bar{\phi}_{i-1}(t)\bar{\vartheta}_{i-1}(t-1)$$
$$- \bar{\phi}_{i+1}(t)\bar{\vartheta}_{i+1}(t-1) - \cdots$$
$$- \bar{\phi}_{M+n+m}(t)\bar{\vartheta}_{M+n+m}(t-1)].$$

Remark 2: In the improved lasso SG algorithm, when the residual error m_i lays between $[-\frac{\lambda}{2}, \frac{\lambda}{2}]$, its corresponding parameter element ϑ_i is equal to zero, which means that the redundant parameter elements are picked out from the vector.

Remark 3: In the improved lasso SG algorithm, each parameter element has its own direction and step-size. Thus it can update the parameters adaptively when the elements have different orders of magnitude.

Remark 4: The improved lasso SG algorithm has faster convergence rates than the traditional SG algorithm for its adaptive property. However, its computational efforts are

H. Zhao *et al.*: Improved Redundant Rule-Based Stochastic Gradient Algorithm for Time-Delayed Models Using Lasso Regression

IEEE *Access*

heavier than those of the traditional SG algorithm because it should compute $M + n + m$ step-sizes in each sampling instant.

B. STEP-SIZE CHOOSING

The step-size is important in the SG algorithm designing, a small step-size leads to a slow convergence rate, while a large one would get divergent results. Next, three step-size designing methods are given.

Method 1:

For simplicity, let

$$n_j(t) = \bar{\phi}_i(t)[y(t) - \bar{\phi}_1(t)\bar{\vartheta}_1(t-1) - \cdots - \bar{\phi}_{M+n+m}(t)\bar{\vartheta}_{M+n+m}(t-1)]. \quad (5)$$

Rewrite the parameter element

$$\bar{\vartheta}_i(t) = \begin{cases} \bar{\vartheta}_i(t-1) + r_i(n_j(t) - \frac{\lambda}{2}), & \text{if } m_i(t) > \frac{\lambda}{2}, \\ 0, & \text{if } m_i(t) \in [-\frac{\lambda}{2}, \frac{\lambda}{2}] \\ \bar{\vartheta}_i(t-1) + r_i(n_j(t) + \frac{\lambda}{2}), & \text{if } m_i < -\frac{\lambda}{2} \end{cases}.$$

Substituting the above equation into the cost function $J(\bar{\boldsymbol{\vartheta}})$ and taking the derivative of $J(r_i)$ with respect to r_i yield

$$r_i(t) = \begin{cases} \dfrac{0.5\lambda - \bar{\phi}_i(t)n_i(t)}{\bar{\phi}_i^2(t)(n_i(t) - \lambda)}, & \text{if } m_i(t) > \frac{\lambda}{2}, \\ \dfrac{\bar{\phi}_i(t)n_i(t) - 0.5\lambda}{\bar{\phi}_i^2(t)(n_i(t) + \lambda)}, & \text{if } m_i < -\frac{\lambda}{2} \end{cases}.$$

Remark 5: Method 1 computes the step-size by solving the derivative function of $J(r_i)$, it is the same as the steepest descent SG algorithm in [26].

Method 2:

Consider that the cost function $J_2(\bar{\boldsymbol{\vartheta}})$ is used to chose the small weights from the parameter vector. When computing the step-size, we can neglect its influence.

Substituting the parameter estimates into the cost function and taking the derivative of $J(r_i)$ with respect to r_i yield

$$r_i(t) = \frac{1}{\bar{\phi}_i^2(t)}.$$

Since a smaller $\bar{\phi}_i(t)$ may yield a larger step-size, the algorithm can be oscillated intensively. To overcome these difficulty, we introduce a small constant ρ. Then the step-size can be written by

$$r_i(t) = \frac{1}{\rho + \bar{\phi}_i^2(t)}.$$

It is noted that when the estimates are closing the true values, the step-size becomes smaller and smaller, another kind of step-size is

$$r_i(t) = \frac{1}{\sum_{j=1}^{t-1} \bar{\phi}_i^2(j)}.$$

In this case, the algorithm is more stable.

Method 3:

Inspired by the intelligent search method, we can use many step-sizes to get a better parameter estimate.

For example, when computing the step-size $r_i(t)$, we first give an interval $[0, r_i^{max}(t)]$ and choose N different step-sizes $r_i^j(t), j = 1, 2, \cdots, N$ from this interval. Substituting these N step-sizes into the parameter estimates gets

$$\bar{\vartheta}_i^j(t) = \begin{cases} \bar{\vartheta}_i(t-1) + r_i^j(n_j(t) - \frac{\lambda}{2}), & \text{if } m_i(t) > \frac{\lambda}{2}, \\ 0, & \text{if } m_i(t) \in [-\frac{\lambda}{2}, \frac{\lambda}{2}] \\ \bar{\vartheta}_i(t-1) + r_i^j(n_j(t) + \frac{\lambda}{2}), & \text{if } m_i < -\frac{\lambda}{2} \end{cases}.$$

Then their corresponding N cost functions are yielded $J(\vartheta_i^j(t))$. The best estimate is the one satisfies

$$\bar{\vartheta}_i(t) = \arg \min_{\bar{\vartheta}_i^j(t)} [J(\bar{\vartheta}_i^j(t)), j = 1, \cdots, N].$$

Remark 6: In method 3, the choice of the upper bound $r_i^{max}(t)$ is challenging, a small upper bound may lead to slow convergence rates, e.g., all the cost functions ensure $J(\bar{\vartheta}_i^j(t)) < J(\bar{\vartheta}_i(t-1))$; on the other hand, a large upper bound may diverge the results, e.g., all the cost functions satisfy $J(\bar{\vartheta}_i^j(t)) > J(\bar{\vartheta}_i(t-1))$.

Remark 7: For a small upper bound, we usually assign $r_{i,new}^{max}(t) = 2r_{i,old}^{max}(t)$; while for a large one, we can perform $r_{i,new}^{max}(t) = \frac{1}{2}r_{i,old}^{max}(t)$.

Remark 8: Since the intelligent search method should compute several cost functions in each iteration, its computational efforts are heavy.

C. PROPERTIES OF THE ALGORITHM

Define

$$\bar{\vartheta}_i(t) - \bar{\vartheta}_i = e_i(t).$$

Subtracting $\bar{\vartheta}_i$ on both sides of the parameter estimates yields

$$e_i(t) = \begin{cases} e_i(t-1) - r_i\bar{\phi}_i(t)[\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)E(t-1) - v(t)] - r_i\frac{\lambda}{2}, \\ \quad \text{if } m_i(t) > \frac{\lambda}{2}, \\ 0, \text{ if } m_i(t) \in [-\frac{\lambda}{2}, \frac{\lambda}{2}] \\ e_i(t-1) - r_i\bar{\phi}_i(t)[\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)E(t-1) - v(t)] + r_i\frac{\lambda}{2}, \\ \quad \text{if } m_i(t) < \frac{\lambda}{2} \end{cases}. \quad (6)$$

For simplicity, we only consider the first equation in (6). The parameter vector error can be written by

$$E(t) = E(t-1) - R(t-1)\bar{\boldsymbol{\phi}}(t)[\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)E(t-1) - v(t)] - R(t-1)\boldsymbol{\lambda}, \quad (7)$$

IEEE *Access*

H. Zhao *et al.*: Improved Redundant Rule-Based Stochastic Gradient Algorithm for Time-Delayed Models Using Lasso Regression

where

$$R(t-1) = diag[r_1(t-1), r_2(t-1), \cdots, r_{M+n+m}(t-1)],$$
$$\boldsymbol{\lambda} = [\frac{\lambda}{2}, \frac{\lambda}{2}, \cdots, \frac{\lambda}{2}]^{M+n+m}.$$

Since $v(t)$ is a Gaussian white noise and is independent on the input and output $\{u(1), \cdots, u(t), y(1), \cdots, y(t-1)\}$, Equation (7) is simplified as

$$E(t) = [I - R(t-1)\bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)]E(t-1) - R(t-1)\boldsymbol{\lambda}. \quad (8)$$

The step-size matrix $R(t-1)$ is usually chosen to ensure

$$\varrho_{max}[I - R(t-1)\bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)] < 1,$$

where $\varrho_{max}[M]$ means the spectral radius of the matrix M.

Taking the expectation on both sides of Equation (7) obtains

$$\boldsymbol{E}[E(t)] = \boldsymbol{E}\{[I - R(t-1)\bar{\boldsymbol{\phi}}(t)\bar{\boldsymbol{\phi}}^{\mathrm{T}}(t)]E(t-1)\} - \boldsymbol{E}[R(t-1)\boldsymbol{\lambda}]$$
$$= R(t-1)\boldsymbol{\lambda} \neq 0,$$

where $0 \in \mathbb{R}^{M+n+m}$ is a zero vector.

Remark 9: The above equation shows that the redundant rule based lasso SG algorithm is a biased algorithm.

V. EXAMPLE

Consider the following time-delayed model,

$$y(t) = 0.12z^{-1}y(t-\tau-1)+[1.2z^{-1}+0.85z^{-2}]u(t) + v(t),$$
$$u(t) \backsim N(0,1), \quad v(t) \backsim N(0,0.1^2).$$

Assume that the time-delay $\tau = 1$, and the upper bound of the time-delay is $M = 4$. Define

$$\bar{\boldsymbol{\vartheta}} = [\bar{\vartheta}_1, \bar{\vartheta}_2, \bar{\vartheta}_3, \bar{\vartheta}_4, \bar{\vartheta}_5, \bar{\vartheta}_6, \bar{\vartheta}_7]^{\mathrm{T}}$$
$$= [0, 0.12, 0, 0, 0, 1.2, 0.85]^{\mathrm{T}},$$
$$\boldsymbol{\phi}(t) = [y(t-1), y(t-2), y(t-3), y(t-4), y(t-5),$$
$$u(t-1), u(t-2)]^{\mathrm{T}}.$$

Apply the redundant rule based SG (RR-SG) algorithm and the redundant rule based lasso SG (RR-LR-SG, $\lambda = 0.8$) algorithm for this time-delayed model, the parameter estimates and their errors are shown in Table 1, the parameter estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ versus t are shown in Figure 1.

To show the role of the constants λ, use the RR-LR-SG algorithm with different λ for the time-delayed model, the parameter estimates and their errors are shown in Table 2.

In addition, we apply the RR-LR-SG algorithm for the considered model with different signal-to-noise ratios (SNR). The parameter estimation errors are shown in Figure 2. Finally, the step-size choosing methods proposed in Section IV are applied for the time-delayed model, the parameter estimation errors are shown in Figure 3.

From this example, the following findings can be obtained:

(1) The parameter estimates by using the SG algorithm can asymptotically converge to a stable point but not the optimal
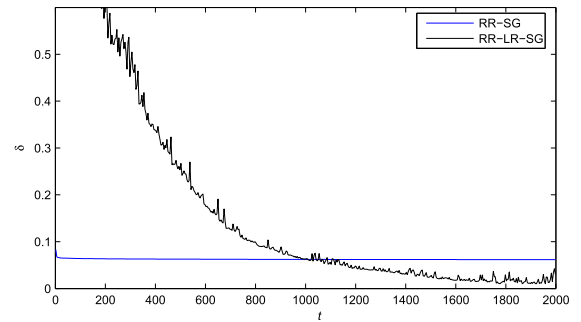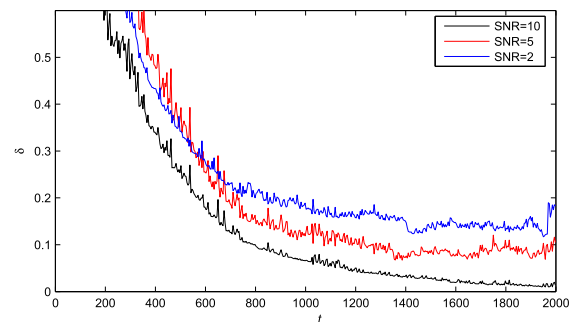

FIGURE 1. The parameter estimation errors δ versus t.

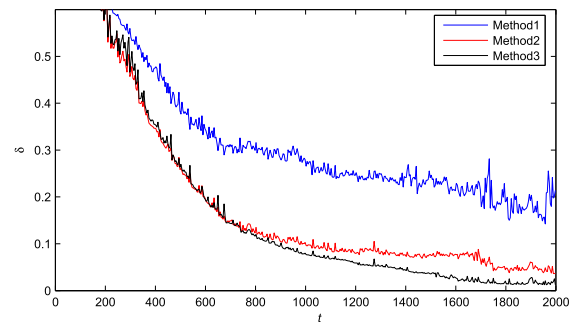**FIGURE 2.** The parameter estimation errors δ versus t.

FIGURE 3. The parameter estimation errors δ versus t.

point that is because the step-sizes are approaching to zero with the increased sampling instant t, as shown in Table 1 and Figure 1.

(2) The parameter estimates by using the RR-LR-SG algorithm are more accurate than those of the SG algorithms, for the reason that the RR-LR-SG algorithm updates the parameter elements one by one with adaptive step-sizes and directions, as shown in Table 1 and Figure 1.

(3) The RR-LR-SG algorithm can pick out the redundant elements from the parameter vector, while the SG algorithm cannot, as shown in Table 1.

(4) A larger λ can lead to more accurate time-delay estimate but less accurate parameter estimates, this is demonstrated in Table 2.

(5) When using the RR-LR-SG algorithm for the proposed time-delayed model, a larger SNR leads to more accurate parameter estimates, see Figure 2.

H. Zhao *et al.*: Improved Redundant Rule-Based Stochastic Gradient Algorithm for Time-Delayed Models Using Lasso Regression

IEEE *Access*

TABLE 1. The parameter estimates and errors.

Algorithm	t	$\bar{\vartheta}_1$	$\bar{\vartheta}_2$	$\bar{\vartheta}_3$	$\bar{\vartheta}_4$	$\bar{\vartheta}_5$	$\bar{\vartheta}_6$	$\bar{\vartheta}_7$	δ (%)
	500	0.033	0.045	0.013	0.004	0.001	1.232	0.875	6.273
RR-SG	1000	0.033	0.045	0.013	0.004	0.001	1.232	0.874	6.214
	1500	0.033	0.046	0.013	0.004	0.001	1.232	0.874	6.185
	2000	0.033	0.046	0.013	0.004	0.001	1.232	0.874	6.166
	500	-0.168	0.072	-0.032	0.048	0.181	0.914	0.733	27.282
RR-LR-SG	1000	-0.006	0.110	-0.008	-0.025	0.073	1.153	0.840	6.217
	1500	0.002	0.119	-0.004	-0.004	0.031	1.187	0.853	2.324
	2000	0.000	0.138	0.000	0.000	0.000	1.184	0.869	1.676
True Values		0.000	0.120	0.000	0.000	0.000	1.200	0.850	

TABLE 2. The parameter estimates and errors (RR-LR-SG).

	t	$\bar{\vartheta}_1$	$\bar{\vartheta}_2$	$\bar{\vartheta}_3$	$\bar{\vartheta}_4$	$\bar{\vartheta}_5$	$\bar{\vartheta}_6$	$\bar{\vartheta}_7$	δ (%)
	500	-0.367	0.011	0.036	0.136	0.272	0.565	0.785	54.556
$\lambda = 0.1$	1000	-0.018	0.058	0.012	-0.036	0.137	1.091	0.906	13.436
	1500	-0.006	0.090	0.005	-0.005	0.051	1.172	0.858	4.506
	2000	0.003	0.126	0.006	0.004	0.004	1.199	0.865	1.226
	500	-0.132	0.066	-0.037	0.042	0.175	0.910	0.728	26.520
$\lambda = 0.5$	1000	-0.020	0.104	-0.002	-0.007	0.071	1.150	0.836	6.230
	1500	0.011	0.111	-0.001	0.000	0.027	1.184	0.842	2.425
	2000	0.000	0.134	0.002	0.004	0.003	1.200	0.867	1.436
	500	-0.168	0.072	-0.032	0.048	0.181	0.914	0.733	27.282
$\lambda = 0.8$	1000	-0.006	0.110	-0.008	-0.025	0.073	1.153	0.840	6.217
	1500	0.002	0.119	-0.004	-0.004	0.031	1.187	0.853	2.324
	2000	0.000	0.138	0.000	0.000	0.000	1.184	0.869	1.676
True Values		0.000	0.120	0.000	0.000	0.000	1.200	0.850	

(6) Figure 3 shows that the intelligent search method (Method 3) has the best step-size, then is Method 2, and Method 1 has the poorest step-size.
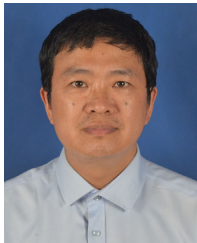
VI. CONCLUSION

An improved redundant rule based lasso regression stochastic gradient algorithm is proposed for time-delayed model in this study. The proposed algorithm has two advantages when compared with the traditional SG algorithm: (1) each element in the parameter vector can be adaptively updated (has different step-sizes and directions); (2) the parameter and time-delay estimates can be obtained simultaneously. The convergence analysis and simulation example are given to show that the proposed algorithm is effective.

Although the proposed algorithm can get the estimates adaptively, it remains some challenging and interesting problems. For example, the choice of the values of the constant λ, the step-size in the improved SG algorithm. These topics need to be further investigated.

REFERENCES

[1] F. Ding, L. Lv, J. Pan, X. Wan, and X.-B. Jin, "Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data," *Int. J. Control, Autom. Syst.*, vol. 18, no. 4, pp. 886–896, Apr. 2020.

[2] J. Chen, F. Ding, Q. Zhu, and Y. Liu, "Interval error correction auxiliary model based gradient iterative algorithms for multirate ARX models," *IEEE Trans. Autom. Control*, vol. 65, no. 10, pp. 4385–4392, Oct. 2020.

[3] H. Xia and F. Chen, "Filtering-based parameter identification methods for multivariable stochastic systems," *Mathematics*, vol. 8, no. 12, p. 2254, Dec. 2020.

[4] F. Ding, J. Pan, A. Alsaedi, and T. Hayat, "Gradient-based iterative parameter estimation algorithms for dynamical systems from observation data," *Mathematics*, vol. 7, no. 5, p. 428, May 2019.

[5] H. Ma, J. Pan, L. Lv, G. Xu, F. Ding, A. Alsaedi, and T. Hayat, "Recursive algorithms for multivariable output-error-like ARMA systems," *Mathematics*, vol. 7, no. 6, p. 558, Jun. 2019.

[6] J. Chen, B. Huang, M. Gan, and C. L. P. Chen, "A novel reduced-order algorithm for rational models based on Arnoldi process and Krylov subspace," *Automatica*, vol. 129, Jul. 2021, Art. no. 109663, doi: 10.1016/j.automatica.2021.109663.

[7] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.

[8] S. Magnusson, C. Enyioha, N. Li, C. Fischione, and V. Tarokh, "Convergence of limited communication gradient methods," *IEEE Trans. Autom. Control*, vol. 63, no. 5, pp. 1356–1371, May 2018.

[9] F. Ding, *System Identification—Multi-Innovation Identification Theory and Methods*. Beijing, China: Science Press, 2016.

[10] H. Xu, F. Ding, and E. Yang, "Three-stage multi-innovation parameter estimation for an exponential autoregressive time-series model with moving average noise by using the data filtering technique," *Int. J. Robust Nonlinear Control*, vol. 31, no. 1, pp. 166–184, Jan. 2021.

[11] J. Ma, W. Xiong, F. Ding, A. Alsaedi, and T. Hayat, "Data filtering based forgetting factor stochastic gradient algorithm for Hammerstein systems with saturation and preload nonlinearities," *J. Franklin Inst.*, vol. 353, no. 16, pp. 4280–4299, Nov. 2016.

[12] J. Ma, Q. Fei, F. Guo, and W. Xiong, "Variational Bayesian iterative estimation algorithm for linear difference equation systems," *Mathematics*, vol. 7, no. 12, p. 1143, Nov. 2019.

[13] A. G. Carlon, B. M. Dia, L. Espath, R. H. Lopez, and R. Tempone, "Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization," *Comput. Methods Appl. Mech. Eng.*, vol. 363, May 2020, Art. no. 112909, doi: 10.1016/j.cma.2020.112909.

[14] Z.-X. Cui, Q. Fan, and C. Jia, "Momentum methods for stochastic optimization over time-varying directed networks," *Signal Process.*, vol. 174, Sep. 2020, Art. no. 107614, doi: 10.1016/j.sigpro.2020.107614.

[15] J. Chen, B. Huang, F. Ding, and Y. Gu, "Variational Bayesian approach for ARX systems with missing observations and varying time-delays," *Automatica*, vol. 94, pp. 194–204, Aug. 2018.

[16] M. J. Moghaddam, H. Mojallali, and M. Teshnehlab, "Recursive identification of multiple-input single-output fractional-order Hammerstein model with time delay," *Appl. Soft Comput.*, vol. 70, pp. 486–500, Sep. 2018.

[17] C. Wang and K. Li, "Aitken-based stochastic gradient algorithm for ARX models with time delay," *Circuits, Syst., Signal Process.*, vol. 38, no. 6, pp. 2863–2876, Jun. 2019.

[18] L. Xie, H. Yang, and B. Huang, "FIR model identification of multirate processes with random delays using EM algorithm," *AIChE J.*, vol. 59, no. 11, pp. 4124–4132, Nov. 2013.

[19] J.-N. Li, Y. Zhang, and Y.-J. Pan, "Mean-square exponential stability and stabilisation of stochastic singular systems with multiple time-varying delays," *Circuits, Syst., Signal Process.*, vol. 34, no. 4, pp. 1187–1210, Apr. 2015.

[20] Y. Zhao, A. Fatehi, and B. Huang, "Robust estimation of ARX models with time varying time delays using variational Bayesian approach," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 532–542, Feb. 2018.

[21] J. Chen, Q. Shen, Y. Liu, and L. Wan, "Expectation maximization identification algorithm for time-delay two-dimensional systems," *J. Franklin Inst.*, vol. 357, no. 14, pp. 9992–10009, Sep. 2020.

[22] J. Ma, J. Chen, W. Xiong, and F. Ding, "Expectation maximization estimation algorithm for Hammerstein models with non-Gaussian noise and random time delay from dual-rate sampled-data," *Digit. Signal Process.*, vol. 73, pp. 135–144, Feb. 2018.

[23] D. Wang, L. Li, Y. Ji, and Y. Yan, "Model recovery for Hammerstein systems using the auxiliary model based orthogonal matching pursuit method," *Appl. Math. Model.*, vol. 54, pp. 537–550, Feb. 2018.

[24] J. Chen, J. Ma, Y. Liu, and F. Ding, "Identification methods for time-delay systems based on the redundant rules," *Signal Process.*, vol. 137, pp. 192–198, Aug. 2017.

[25] X. Zhang, Q. Liu, F. Ding, A. Alsaedi, and T. Hayat, "Recursive identification of bilinear time-delay systems through the redundant rule," *J. Franklin Inst.*, vol. 357, no. 1, pp. 726–747, Jan. 2020.

[26] T. Söderström and P. Stoica, *Systen Identification*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

LIXIN LV received the B.Sc. and M.Sc. degrees from the Xi'an University of Science and Technology, Xi'an, China, in 1988 and 1990, respectively. He is currently an Associate Professor with the School of Internet of Things Engineering, Wuxi Vocational College of Science and Technology, Wuxi, China. His current research interests include system identification, adaptive control, and artificial intelligence.

HANGTAO ZHAO received the B.Sc. degree from Nanjing Normal University, Nanjing, China, in 1998, and the M.Sc. degree from the School of Internet of Things Engineering, Jiangnan University, Wuxi, China, in 2014. He is currently an Associate Professor with the School of Internet of Things Engineering, Wuxi Vocational College of Science and Technology, Wuxi. His current research interests include system identification and machine learning.

YUEJIANG JI received the B.Sc. and M.Sc. degrees from Jiangnan University, Wuxi, China, in 2004 and 2007, respectively. He is currently an Associate Professor with the Wuxi Vocational College of Science and Technology, Wuxi. His current research interests include machine learning, process control, and artificial intelligence.

• • •