Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(7): 16898–16926.
DOI: 10.3934/math.2025760
Received: 10 May 2025
Revised: 06 July 2025
Accepted: 15 July 2025
Published: 29 July 2025

Research article

Synchronous parameter and state estimation for bilinear state-space systems with censored measurements

Xuehai Wang* **and Fang Zhu**

School of Mathematics and Statistics, Xinyang Normal University, Xinyang 464000, China

* **Correspondence:** Email: xuehaiwang735@163.com.

Abstract: The censored measurements are unavoidably encountered in practical scenarios due to sudden changes of circumstances and physical constraints, and the traditional parameter and state estimation methods may result in obvious estimation bias and serious performance degradation when the collected measurements contain censored observations. This paper concerns the synchronous parameter and state estimation for the bilinear state-space systems with censored measurements. By compensating the estimation bias from the censored measurements and constructing a novel criterion function, a censored regression-based gradient method is presented for synchronously estimating the system parameters and states. Moreover, a censored regression-based multi-innovation gradient method is derived to enhance the estimation performance. Theoretical analysis reveals that the convergence of the parameter estimates can be guaranteed under the persistent excitation condition. The simulation examples exhibit that the proposed method performs significantly better than the conventional auxiliary model-based least mean square method.

Keywords: parameter estimation; state estimation; bilinear state-space system; bias compensation; censored measurement
Mathematics Subject Classification: 93E11, 93E12, 93E23, 93E24

1. Introduction

System modeling and identification are important in system analysis, controller design, and fault diagnosis, and parameter estimation is the prerequisite for implementing control strategies of practical systems [1, 2]. Since most of the control systems can be regarded as the nonlinear systems, the nonlinear system identification has attracted significant focus from research communities and engineers [3]. In recent years, various of nonlinear system structures have been reported, such as the rational models [4], Volterra series models [5, 6], block-oriented nonlinear models [7, 8], neural networks [9], and bilinear models [10]. As a special class, the bilinear model contains the coupling terms of the control variables and the system states [11, 12], and it has relatively simple model structures but can approximately describe a wide range of complex dynamic processes such as chemical, biological, and robotic processes [13, 14].

When the system states are available, the conventional linear system identification methods such as the recursive least squares algorithm and gradient descent algorithm can be applied to the bilinear system identification [15, 16]. When the states are not available, the identification of the bilinear system is more challenging due to the existing of the unknown bilinear product term. A feasible choice is to transform the bilinear system into the input-output model by means of the state variables elimination scheme and then to identify the parameters of the bilinear system [17, 18]. Another choice is to synchronously estimate the parameters and states of the bilinear systems based on the state-space representation [19, 20]. In this aspect, Liu et al. used the modified Kalman filter to estimate the states and developed a gradient iterative algorithm for the bilinear systems with colored noise [21]. The aforementioned works were accomplished based on the assumption that the output data can be directly collected.

It is well-known that the measurements play a vital role for the parameter estimation and adaptive filtering. In many practical scenarios, especially those using low cost commercial off-the-shelf sensors, the values of the measurements may exceed the limit of the recording devices such that a proportion of sampling data cannot be recorded by the devices [22–24]. This situation may also occur due to physical constraints such as sensor saturation, limit-of-detection effects, and occlusion region [25–27]. This special type of measurement nonlinearity is referred to as the censored measurement. If this phenomenon is not properly handled, the performance of the system parameter and state estimation will be inevitably degraded [28–30]. To deal with such a problem, the censored outputs can be seen as the missing data, and then interval-varying identification algorithms and maximum likelihood algorithms can be used [31, 32]. The problems of these methods lie in that when the output data are censored, the update of the parameter estimation may be halted and a low estimation accuracy may be resulted. In addition, the censored mechanism may be regarded as a special class of output nonlinearity and the Wiener nonlinear system identification methods can be applied [33]. However, the uncertainty of the stochastic noise may induce that some of measurements are improperly censored, which leads to the biased estimates.

In recent years, considerable efforts have been devoted to overcoming the aforementioned drawbacks. A maximum likelihood method was applied to estimate the parameters of Weibull generalized exponential distribution with censoring mechanisms in [34]. A recursive state estimation method has been developed for the linear systems with random transmission loss and censored measurements, where the occurrence probability of censoring was utilized to maintain the unbiased estimate in [35]. Subsequently, an adaptive Tobit Kalman filter is presented for tracking the maneuvering target, where the censoring probability can be dynamically computed in [36]. A complex domain adaptive filter is proposed for the widely linear systems with censored measurements based on the expectation maximization scheme in [37].

To the best of the authors' knowledge, although the identification of the systems with censored measurements has been discussed in open literatures, the joint parameter and state estimation issue of the bilinear state-space systems with censored measurements are rarely reported. This paper attempts to challenge this issue when the states are not available. The difficulties are that not only the bilinear state-space system contains the unknown states and bilinear product terms, but also the observed system output is censored. The main contributions of this paper are threefold as follows.

- Construct a criterion function considering the estimation bias caused by the censored measurements to obtain the consistent parameter estimation of the bilinear systems.
- Present a censored regression-based gradient method to simultaneously estimate the system parameters and states of the bilinear state-space systems.
- Develop a censored regression-based multi-innovation gradient method for the bilinear state-space systems with censored measurements to enhance the estimation performance.

The rest of the paper is organized as follows. Section 2 gives the system description and the identification model. Section 3 presents the censored regression-based gradient (CRG) method for the bilinear state-space systems with censored measurements based on the bias compensation. Section 4 derives the censored regression-based multi-innovation gradient (CRMIG) method. Section 5 provides a numerical example to validate the effectiveness of the proposed methods. Section 6 gives some conclusions.

2. Problem description

Consider the following bilinear state-space system in the observability canonical form:

$$\boldsymbol{x}(s+1) = \boldsymbol{G}\boldsymbol{x}(s) + \boldsymbol{F}\boldsymbol{x}(s)u(s) + \boldsymbol{h}u(s), \tag{2.1}$$

$$y^*(s) = \boldsymbol{c}\boldsymbol{x}(s) + v(s), \tag{2.2}$$

where s denotes the instant, $\boldsymbol{x}(s) := [x_1(s), x_2(s), \cdots, x_M(s)]^{\mathrm{T}} \in \mathbb{R}^M$, $u(s) \in \mathbb{R}$, and $y^*(s) \in \mathbb{R}$ are the system state vector, input, and output, M is the order of the bilinear system, $v(s) \in \mathbb{R}$ is the stochastic white noise with zero mean and bounded variance σ^2, and $\boldsymbol{G} \in \mathbb{R}^{M\times M}$, $\boldsymbol{F} \in \mathbb{R}^{M\times M}$, $\boldsymbol{h} \in \mathbb{R}^M$, and $\boldsymbol{c} \in \mathbb{R}^{1\times M}$ denote the parameter matrices and vectors:

$$\boldsymbol{G} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -g_M & -g_{M-1} & -g_{M-2} & \cdots & -g_1 \end{bmatrix} \in \mathbb{R}^{M\times M},$$

$$\boldsymbol{F} := \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \vdots \\ \boldsymbol{f}_M \end{bmatrix} \in \mathbb{R}^{M\times M}, \boldsymbol{f}_i := [f_{i1}, f_{i2}, \cdots, f_{iM}] \in \mathbb{R}^{1\times M}, i = 1, 2, \cdots, M,$$

$$\boldsymbol{h} := [h_1, h_2, \cdots, h_M]^{\mathrm{T}} \in \mathbb{R}^M,$$

$$\boldsymbol{c} := [1, 0, \cdots, 0, 0] \in \mathbb{R}^{1\times M}.$$

Due to various reasons such as unreliable communication link or the saturation characteristics of the sensor, only a proportion of output data can be available. In other words, the output data within the predetermined range can be accurately measured, and other output data are distorted, i.e., the system output $y^*(s)$ is censored by

$$y(s) = \begin{cases} y^*(s), & y^*(s) > 0 \\ 0, & y^*(s) \leqslant 0 \end{cases} \tag{2.3}$$

where $y(s)$ represents the measured output. Equation (2.3) is the mathematical representation of the censored measurement and is often referred to as the left censored regression model.

Suppose that the system order M is known, and $u(s) = 0$, $\boldsymbol{x}(s) = \mathbf{0}$, and $y(s) = 0$ for $s \leqslant 0$. Substituting the expressions of $\boldsymbol{G}$, $\boldsymbol{F}$, and $\boldsymbol{h}$ into (2.1) gives

$$\begin{bmatrix} x_1(s+1) \\ x_2(s+1) \\ \vdots \\ x_{M-1}(s+1) \\ x_M(s+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -g_M & -g_{M-1} & -g_{M-2} & \cdots & -g_1 \end{bmatrix} \begin{bmatrix} x_1(s) \\ x_2(s) \\ \vdots \\ x_{M-1}(s) \\ x_M(s) \end{bmatrix}$$
$$+ \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \vdots \\ \boldsymbol{f}_{M-1} \\ \boldsymbol{f}_M \end{bmatrix} \boldsymbol{x}(s)u(s) + \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_{M-1} \\ h_M \end{bmatrix} u(s).$$

Thus, we have

$$\begin{aligned} x_i(s+1) &= x_{i+1}(s) + \boldsymbol{f}_i \boldsymbol{x}(s)u(s) + h_i u(s),\ i = 1, 2, \cdots, M-1, \\ x_M(s+1) &= -g_M x_1(s) - g_{M-1}x_2(s) - \cdots - g_1 x_M(s) + \boldsymbol{f}_M \boldsymbol{x}(s)u(s) + h_M u(s). \end{aligned}$$

Multiplying both sides of the i-th equation by $z^{-i}(z^{-i}y(s) = y(s-i), i = 1, 2, \cdots, M)$ and adding all these expressions yields

$$x_1(s) = -\sum_{m=1}^{M} g_m x_{M-m+1}(s-n) + \sum_{m=1}^{M} \boldsymbol{f}_i \boldsymbol{x}(s-m)u(s-m) + \sum_{m=1}^{M} h_m u(s-m). \tag{2.4}$$

Define

$$\begin{aligned} \boldsymbol{\psi}(s) &:= [-x_M(s-M), \cdots, -x_1(s-M), u(s-1)\boldsymbol{x}^{\mathrm{T}}(s-1), \cdots, u(s-M)\boldsymbol{x}^{\mathrm{T}}(s-M), u(s-1), \cdots, u(s-M)]^{\mathrm{T}} \\ &\in \mathbb{R}^{M^2+2M}, \\ \boldsymbol{\theta} &:= [g_1, g_2, \cdots, g_M, \boldsymbol{f}_1^{\mathrm{T}}, \boldsymbol{f}_2^{\mathrm{T}}, \cdots, \boldsymbol{f}_M^{\mathrm{T}}, h_1, h_2, \cdots, h_M]^{\mathrm{T}} \in \mathbb{R}^{M^2+2M}. \end{aligned}$$

Substituting (2.4) into (2.2) yields

$$\begin{aligned} y^*(s) &= x_1(s) + v(s) \\ &= -\sum_{m=1}^{M} g_m x_{M-m+1}(s-M) + \sum_{m=1}^{M} \boldsymbol{f}_l \boldsymbol{x}(s-m)u(s-m) + \sum_{m=1}^{M} h_m u(s-m) + v(s) \\ &= \boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + v(s). \end{aligned} \tag{2.5}$$

Thus, the censored output in (2.3) can be written as

$$y(s)=\begin{cases}\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}+v(s), & y^*(s)>0\\ 0. & y^*(s)\leqslant 0\end{cases}\tag{2.6}$$

Note that when $y^*(s)\leqslant 0$, the system output $y^*(s)$ is censored and the measured output is $y(s)=0$. To identify the bilinear states-space system with censored measurements, a simple way is to regard the censored data as the missing data and identify the system only from the uncensored measurements. Assume that the system output $y^*(s)$ is not censored at the sampling instants $s=s_1,s_2,\cdots,s_l,\cdots$, where s_l represents the l-th instant that meets $y^*(s)>0$ and the integer sequence $\{s_l:\ l=0,1,2,\cdots,\}$ satisfies:

$$0=s_0<s_1<s_2<\cdots<s_l<\cdots,$$

and $s_l^*=s_l-s_{l-1}\geqslant 1$. Using (2.3), it gives $y(s_l)=y^*(s_l)$ at the sampling instant s_l. Replacing s with s_l in (2.5) gives

$$y(s_l)=\boldsymbol{\psi}^{\mathrm{T}}(s_l)\boldsymbol{\theta}+v(s_l).$$

Define the cost function

$$J_1(\boldsymbol{\theta}):=\|y(s_l)-\boldsymbol{\psi}^{\mathrm{T}}(s_l)\boldsymbol{\theta}\|^2.$$

Minimizing $J_1(\boldsymbol{\theta})$ and substituting the unknown $\boldsymbol{x}(s)$ in $\boldsymbol{\psi}(s_l)$ by its estimate $\hat{\boldsymbol{x}}(s)$ gives the following recursive relations:

$$\hat{\boldsymbol{\theta}}(s_l)=\hat{\boldsymbol{\theta}}(s_{l-1})+\frac{\hat{\boldsymbol{\psi}}(s_l)}{1+\|\hat{\boldsymbol{\psi}}(s_l)\|^2}[y(s_l)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s_l)\hat{\boldsymbol{\theta}}(s_{l-1})],\tag{2.7}$$

$$\hat{\boldsymbol{\theta}}(s)=\hat{\boldsymbol{\theta}}(s_l),\ t\in S_l:=\{s_l+1,s_l+2,\cdots,s_{l+1}-1\},\tag{2.8}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}(s_l)=&[-\hat{x}_M(s_l-M),\cdots,-\hat{x}_1(s_l-M),u(s_l-1)\hat{\boldsymbol{x}}^{\mathrm{T}}(s_l-1),\\ &\cdots,u(s_l-M)\hat{\boldsymbol{x}}^{\mathrm{T}}(s_l-M),u(s_l-1),\cdots,u(s_l-M)]^{\mathrm{T}},\end{aligned}\tag{2.9}$$

$$\hat{\boldsymbol{\theta}}(s)=[\hat{g}_1(s),\hat{g}_2(s),\cdots,\hat{g}_M(s),\hat{\boldsymbol{f}}_1^{\mathrm{T}}(s),\hat{\boldsymbol{f}}_2^{\mathrm{T}}(s),\cdots,\hat{\boldsymbol{f}}_M^{\mathrm{T}}(s),\hat{h}_1(s),\hat{h}_2(s),\cdots,\hat{h}_M(s)]^{\mathrm{T}}.\tag{2.10}$$

From (2.1), the unknown state $\boldsymbol{x}(s+1)$ can be estimated by

$$\hat{\boldsymbol{x}}(s+1)=\hat{\boldsymbol{G}}(s)\hat{\boldsymbol{x}}(s)+\hat{\boldsymbol{F}}(s)\hat{\boldsymbol{x}}(s)u(s)+\hat{\boldsymbol{h}}(s)u(s),\tag{2.11}$$

$$\hat{\boldsymbol{G}}(s)=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 1\\ -\hat{g}_M(s) & -\hat{g}_{M-1}(s) & -\hat{g}_{M-2}(s) & \cdots & -\hat{g}_1(s)\end{bmatrix},\tag{2.12}$$

$$\hat{\boldsymbol{F}}(s)=\begin{bmatrix}\hat{\boldsymbol{f}}_1(s)\\ \hat{\boldsymbol{f}}_2(s)\\ \vdots\\ \hat{\boldsymbol{f}}_M(s)\end{bmatrix},\quad \hat{\boldsymbol{h}}(s)=\begin{bmatrix}\hat{h}_1(s)\\ \hat{h}_2(s)\\ \vdots\\ \hat{h}_M(s)\end{bmatrix}.\tag{2.13}$$

Equations (2.7)–(2.13) constitute the auxiliary model-based least mean square (AM-LMS) method for the bilinear state-space system in (2.1)–(2.3). Although the AM-LMS method can identify the bilinear state-space system with censored measurements, it can be seen that the state $\boldsymbol{x}(s)$ is estimated using the parameter estimate $\hat{\boldsymbol{\theta}}(s)$, which keeps unchanged over the interval $[s_l+1, s_{l+1}-1]$. It means that when a large amount of measurements are censored, the outputs may remain unchanged during a long period. Thus the parameter estimation update is halted, which results in that both the state estimates and the parameter estimates may have the low estimation accuracy. To improve the estimation performance, the following derives a censored regression-based gradient method for the bilinear state-space system with censored measurements.

3. Censored regression-based gradient method

To derive the CRG method, the following lemma is introduced.

Lemma 1. Assume that ξ is a random variable that follows the Gaussian, uniform or Laplace distribution with zero-mean and variance σ^2, then we have

$$\mathrm{E}[\xi|\xi > \tau] = \frac{\varphi(-\tau)}{\Phi(-\tau)} =: \Omega(-\tau),$$

where $\mathrm{E}[\cdot]$ represents the expectation operator, $\Phi(\tau)$ is the cumulative distribution function (cdf) of the random variable ξ, $\varphi(\tau) := \int_{\tau}^{+\infty} \xi\phi(\xi)\mathrm{d}\xi$, $\phi(\xi)$ is the probability density function (pdf) and $\Omega(\tau)$ is called the inverse Mills ratio [38].

It can be seen from (2.3) that the uncensored output is $y^*(s) > 0$ and $y(s) = y^*(s)$ when the censored output is $y(s) > 0$. Using Lemma 1, we have

$$\begin{aligned}
\mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}, y(s) > 0] &= \mathrm{E}[y^*(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}, y^*(s) > 0] \\
&= \mathrm{E}[\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + v(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}, \boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + v(s) > 0] \\
&= \boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \mathrm{E}[v(s)|v(s) > -\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}] \\
&= \boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \Omega\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right).
\end{aligned} \tag{3.1}$$

Equation (3.1) indicates that a bias exists between the actual output $\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}$ and the expectation of the measured output $y(s)$ under the condition $y(s) > 0$. Note that $v(s)$ is a zero-mean noise with variance σ^2, thus, we have

$$\begin{aligned}
\Pr(y(s) > 0) &= \Pr(y^*(s) > 0) = \Pr(v(s) > -\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}) \\
&= 1 - \Pr\left(v(s) \leqslant -\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right) \\
&= 1 - \Phi\left(-\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right) \\
&= \Phi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right),
\end{aligned}$$

where $\Pr(\cdot)$ represents the probability function. Using (3.1), we have

$$\begin{aligned}
\mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}] &= \Pr(y(s) > 0)\mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}, y(s) > 0] + \Pr(y(s) = 0)\mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}, y(s) = 0] \\
&= \Phi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \varphi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right).
\end{aligned} \tag{3.2}$$

Using (3.2) and the probit regression, the censored output can be expressed as

$$y(s) = \Phi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \varphi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right) + v(s). \tag{3.3}$$

Table 1 lists the pdf $\phi(\tau)$, the cdf $\Phi(\tau)$, and $\varphi(\tau)$ of three different background noises (Gaussian, Uniform, and Laplace noise) with zero mean, which is often used in censored regression.

Table 1. The pdf $\phi(\tau)$, cdf $\Phi(\tau)$, and $\varphi(\tau)$ of three different background noises with zero mean.

Background noises	$\phi(\tau)$	$\Phi(\tau)$	$\varphi(\tau)$
Gaussian noise	$\phi(\tau)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{\tau^2}{2\sigma^2}}$	$\Phi(\tau)=\int_{-\infty}^{\tau}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{\tau^2}{2\sigma^2}}\mathrm{d}\tau$	$\varphi(\tau)=\frac{\sigma}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\tau^2}{2\sigma^2}}$
Uniform noise	$\phi(\tau)=\begin{cases}\frac{1}{2a}, & -a<\tau<a\\ 0. & \text{otherwise}\end{cases}$	$\Phi(\tau)=\begin{cases}0, & \tau\leqslant -a\\ \frac{\tau+a}{2a}, & -a<\tau<\tau\\ 1. & \tau\geqslant a\end{cases}$	$\varphi(\tau)=\begin{cases}\frac{a^2-\tau^2}{4a}, & -a<\tau<a\\ 0. & \text{otherwise}\end{cases}$
Laplace noise	$\phi(\tau)=\frac{1}{2\lambda}\mathrm{e}^{-\frac{\lvert\tau\rvert}{\lambda}}$	$\Phi(\tau)=\begin{cases}\frac{1}{2}\mathrm{e}^{\frac{\tau}{\lambda}}, & \tau<0\\ 1-\frac{1}{2}\mathrm{e}^{-\frac{\tau}{\lambda}}. & \tau\geqslant 0\end{cases}$	$\varphi(\tau)=\begin{cases}\frac{1}{2}(\lambda-\tau)\mathrm{e}^{\frac{\tau}{\lambda}}, & \tau<0\\ \frac{1}{2}(\lambda+\tau)\mathrm{e}^{-\frac{\tau}{\lambda}}. & \tau\geqslant 0\end{cases}$

When $v(s)$ is a stochastic noise which follows one of the three distributions described in Table 1, we have

$$\varphi'(\tau)=-\tau\phi(\tau).$$

Define

$$\begin{aligned}J(\boldsymbol{\theta})&=\frac{1}{2}[y(s)-\mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}]]^2\\&=\frac{1}{2}\left[y(s)-\Phi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}-\varphi\left(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right)\right]^2.\end{aligned}\tag{3.4}$$

Then, we have

$$\begin{aligned}\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}&=-[\phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}+\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})-\phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}]\boldsymbol{\psi}(s)e(s,\boldsymbol{\theta})\\&=-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})\boldsymbol{\psi}(s)e(s,\boldsymbol{\theta}),\end{aligned}$$

where

$$e(s,\boldsymbol{\theta}):=y(s)-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}-\varphi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})\in\mathbb{R}.$$

Using the negative gradient search gives

$$\begin{aligned}\hat{\boldsymbol{\theta}}(s)&=\hat{\boldsymbol{\theta}}(s-1)-\mu(s)\frac{\partial J(\hat{\boldsymbol{\theta}}(s-1))}{\partial\boldsymbol{\theta}}\\&=\hat{\boldsymbol{\theta}}(s-1)+\mu(s)\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}(s)e(s,\hat{\boldsymbol{\theta}}(s-1)),\end{aligned}\tag{3.5}$$

$$e(s,\hat{\boldsymbol{\theta}}(s-1))=y(s)-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)-\varphi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)),\tag{3.6}$$

where $\mu(s)>0$ is the step-size. For the recursive algorithms, it is desired that the parameter estimates can catch their true values as s increases, thus, the parameter estimate $\hat{\boldsymbol{\theta}}(s-1)$ at the instant $s-1$ is generally close to $\hat{\boldsymbol{\theta}}(s)$ at the adjacent instant s. Thus, $J(\hat{\boldsymbol{\theta}}(s))$ in (3.4) can be approximated by

$$J(\hat{\boldsymbol{\theta}}(s))\approx\frac{1}{2}[y(s)-\varphi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s)]^2.\tag{3.7}$$

Inserting (3.5) into (3.7) yields

$$\begin{aligned}J(\hat{\boldsymbol{\theta}}(s)) &= \frac{1}{2}\{y(s)-\varphi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}^{\mathrm{T}}(s)\\ &\quad\times[\hat{\boldsymbol{\theta}}(s-1)+\mu(s)\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}(s)e(s,\hat{\boldsymbol{\theta}}(s-1))]\}^2\\ &= \frac{1}{2}\{y(s)-\varphi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)\\ &\quad-\mu(s)\Phi^2(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\psi}(s)e(s,\hat{\boldsymbol{\theta}}(s-1))]\}^2\\ &= \frac{1}{2}\{[1-\mu(s)\Phi^2(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\|\boldsymbol{\psi}(s)\|^2]e(s,\hat{\boldsymbol{\theta}}(s-1))\}^2.\end{aligned}$$

Thus, one can obtain the optimal step-size $\mu(s)$ by

$$\mu(s)=\frac{1}{\Phi^2(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\|\boldsymbol{\psi}(s)\|^2}.$$

To avoid dividing by zero, let $\mu(s):=\frac{1}{r(s)}\in\mathbb{R}$ and $r(s)$ can be computed by

$$r(s)=r(s-1)+\Phi^2(\boldsymbol{\psi}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\|\boldsymbol{\psi}(s)\|^2,\ r(0)=1. \tag{3.8}$$

Therefore, one can summarize the following recursive relations:

$$\hat{\boldsymbol{\theta}}(s)=\hat{\boldsymbol{\theta}}(s-1)+\frac{1}{r(s)}\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\hat{\boldsymbol{\psi}}(s)e(s,\hat{\boldsymbol{\theta}}(s-1)), \tag{3.9}$$

$$r(s)=r(s-1)+\Phi^2(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\|\hat{\boldsymbol{\psi}}(s)\|^2, \tag{3.10}$$

$$e(s,\hat{\boldsymbol{\theta}}(s-1))=y(s)-\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)-\varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)), \tag{3.11}$$

$$\begin{aligned}\hat{\boldsymbol{\psi}}(s)=[&-\hat{x}_M(s-M),\cdots,-\hat{x}_1(s-M),u(s-1)\hat{\boldsymbol{x}}^{\mathrm{T}}(s-1),\cdots,u(s-M)\hat{\boldsymbol{x}}^{\mathrm{T}}(s-M),\\ &u(s-1),\cdots,u(s-M)]^{\mathrm{T}},\end{aligned} \tag{3.12}$$

$$\hat{\boldsymbol{\theta}}(s)=[\hat{g}_1(s),\hat{g}_2(s),\cdots,\hat{g}_M(s),\hat{\boldsymbol{f}}_1^{\mathrm{T}}(s),\hat{\boldsymbol{f}}_2^{\mathrm{T}}(s),\cdots,\hat{\boldsymbol{f}}_M^{\mathrm{T}}(s),\hat{h}_1(s),\hat{h}_2(s),\cdots,\hat{h}_M(s)]^{\mathrm{T}}. \tag{3.13}$$

The system state $\boldsymbol{x}(s)$ is estimated using the state estimator in (2.11)–(2.13). By implementing the interactive estimation of the parameters and states, Eqs (2.11)–(2.13) and (3.9)–(3.13) form the CRG method for the bilinear state-space systems with left censored measurements.

Remark 1. The AM-LMS method in (2.7)–(2.13) updates the parameter estimates only at the instants that the output data are not censored. Compared with the AM-LMS method, the CRG method in (2.11)–(2.13) and (3.9)–(3.13) compensates the bias between the actual output $\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}$ and the measured output $y(s)$ and updates the parameter estimates at every instant by using the prior knowledge of the measurement noise; thus, a higher parameter and state estimation accuracy can be achieved.

Theorem 1. For the CRG method in (2.11)–(2.13) and (3.9)–(3.13), suppose that $v(s)$ is a stochastic noise that follows the Gaussian, Uniform, or Laplace distribution and meets

$$\text{(A1)}\quad \mathrm{E}[v(s)]=0,\quad \mathrm{E}[v^2(s)]=\sigma^2.$$

There exist an integer N and a constant $\alpha>0$ irrelevant to s such that the following persistent excitation condition holds:

$$\text{(A2)}\quad \sum_{m=0}^{N-1}\frac{\hat{\boldsymbol{\psi}}(s+m)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+m)}{r(s+m)}\geqslant\alpha\boldsymbol{I},\ \text{a.s.}$$

Then, $\hat{\boldsymbol{\theta}}(s)$ can converge to its true value $\boldsymbol{\theta}$ in the mean-square sense.

Proof. Let $\Phi(s-1) := \Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)) \in \mathbb{R}$, $\Phi(0) := \Phi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}) \in \mathbb{R}$, $\varphi(s-1) := \varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)) \in \mathbb{R}$, $\varphi(0) := \varphi(\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}) \in \mathbb{R}$, $\tilde{\Phi}(s-1) := \Phi(s-1) - \Phi(0) \in \mathbb{R}$, and $\tilde{\varphi}(s-1) := \varphi(s-1) - \varphi(0) \in \mathbb{R}$. Equation (3.3) can be written as

$$\begin{aligned} y(s) &= \Phi(0)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \varphi(0) + v(s) \\ &= \Phi(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \varphi(s-1) + \kappa(s), \end{aligned} \tag{3.14}$$

where

$$\kappa(s) := -\tilde{\Phi}(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} - \tilde{\varphi}(s-1) + v(s) \in \mathbb{R}.$$

Let $\tilde{\boldsymbol{\theta}}(s) := \hat{\boldsymbol{\theta}}(s) - \boldsymbol{\theta} \in \mathbb{R}^{n_0}$. Subtracting $\boldsymbol{\theta}$ on both sides of (3.9) and using (3.11) and (4.12) yields

$$\begin{aligned} \tilde{\boldsymbol{\theta}}(s) &= \tilde{\boldsymbol{\theta}}(s-1) + \frac{1}{r(s)}\Phi(s-1)\hat{\boldsymbol{\psi}}(s)[\Phi(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} + \kappa(s) - \Phi(s-1)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)] \\ &=: \tilde{\boldsymbol{\theta}}(s-1) + \frac{1}{r(s)}\Phi(s-1)\hat{\boldsymbol{\psi}}(s)[-\Phi(s-1)\Delta(s) + \Phi(s-1)\zeta(s) + \kappa(s)], \end{aligned} \tag{3.15}$$

where

$$\Delta(s) := \hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\tilde{\boldsymbol{\theta}}(s-1) \in \mathbb{R}, \tag{3.16}$$

$$\zeta(s) := [\boldsymbol{\psi}(s) - \hat{\boldsymbol{\psi}}(s)]^{\mathrm{T}}\boldsymbol{\theta} \in \mathbb{R}. \tag{3.17}$$

Squaring (3.15) gives

$$\begin{aligned} \|\tilde{\boldsymbol{\theta}}(s)\|^2 &= \|\tilde{\boldsymbol{\theta}}(s-1)\|^2 + \frac{2}{r(s)}\Phi(s-1)\Delta(s)[-\Phi(s-1)\Delta(s) + \Phi(s-1)\zeta(s) + \kappa(s)] \\ &\quad + \frac{1}{r^2(s)}\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2\{\Phi^2(s-1)\Delta^2(s) - 2\Phi(s-1)\Delta(s)[\Phi(s-1)\zeta(s) + \kappa(s)] \\ &\quad + [\Phi(s-1)\zeta(s) + \kappa(s)]^2\}. \end{aligned} \tag{3.18}$$

Define $W(s) := \|\tilde{\boldsymbol{\theta}}(s)\|^2 \in \mathbb{R}$. Equation (3.10) means $r(s-1) = r(s) - \Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2 \leqslant r(s)$. Using (3.18) yields

$$\begin{aligned} W(s) &= W(s-1) - \left[\frac{2}{r(s)} - \frac{\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2}{r^2(s)}\right]\Phi^2(s-1)\Delta^2(s) \\ &\quad + 2\left[\frac{1}{r(s)} - \frac{\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2}{r^2(s)}\right]\Phi(s-1)\Delta(s)[\Phi(s-1)\zeta(s) + \kappa(s)] \\ &\quad + \frac{1}{r^2(s)}\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2[\Phi^2(s-1)\zeta^2(s) + 2\Phi(s-1)\zeta(s)\kappa(s) + \kappa^2(s)] \\ &= W(s-1) - \frac{1}{r(s)}\Phi^2(s-1)\Delta^2(s) \\ &\quad - \frac{r(s-1)}{r^2(s)}\{\Phi^2(s-1)[\Delta(s) - \zeta(s)]^2 - \Phi^2(s-1)\zeta^2(s) - 2\Phi(s-1)\Delta(s)\kappa(s)\} \\ &\quad + \frac{1}{r^2(s)}\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2[\Phi^2(s-1)\zeta^2(s) + 2\Phi(s-1)\zeta(s)\kappa(s) + \kappa^2(s)]. \end{aligned} \tag{3.19}$$

Let $\phi(s-1) := \phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))$. Using the Taylor formula, we have

$$\begin{aligned}\tilde{\Phi}(s-1) &\approx \Phi'(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))[\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)-\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}]\\ &= \phi(s-1)(\Delta(s)-\zeta(s)),\\ \tilde{\varphi}(s-1) &\approx -\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)\phi(s-1)(\Delta(s)-\zeta(s)).\end{aligned}$$

Thus, $\kappa(s)$ can be expressed as

$$\begin{aligned}\kappa(s) &= -\tilde{\Phi}(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}-\tilde{\varphi}(s-1)+v(s)\\ &= \phi(s-1)[\Delta(s)-\zeta(s)]^2+v(s).\end{aligned}\tag{3.20}$$

Inserting (3.20) into (3.19) yields

$$\begin{aligned}W(s) &= W(s-1)-\frac{1}{r(s)}\Delta^2(s)-\frac{\Phi^2(s-1)-1}{r(s)}\Delta^2(s)-\frac{r(s-1)}{r^2(s)}\Phi^2(s-1)[\Delta(s)-\zeta(s)]^2\\ &\quad+\frac{r(s-1)}{r^2(s)}\left\{\Phi^2(s-1)\zeta^2(s)+2\Phi(s-1)\Delta(s)[\phi(s-1)[\Delta(s)-\zeta(s)]^2+v(s)]\right\}\\ &\quad+\frac{1}{r^2(s)}\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2\{\Phi^2(s-1)\zeta^2(s)+2\Phi(s-1)\zeta(s)[\phi(s-1)[\Delta(s)-\zeta(s)]^2+v(s)]\\ &\quad+\phi^2(s-1)[\Delta(s)-\zeta(s)]^4+2\phi(s-1)[\Delta(s)-\zeta(s)]^2v(s)+v^2(s)\}.\end{aligned}\tag{3.21}$$

Note that $\Phi(s-1)\in(-1,1)$ and $\varphi(s-1)\in(-\sqrt{2\pi},\sqrt{2\pi})$. Assume that $\zeta(s)$ is bounded with $\zeta^2(s)\leqslant\varepsilon$ for given $\varepsilon<\infty$. Since $\Delta(s)$, $\hat{\boldsymbol{\psi}}(s)$, $\zeta(s)$, and $r(s)$ are uncorrelated with $v(s)$, noting that (A1) and $r(s)\to\infty$, and taking the expectation (3.21) yields

$$\mathrm{E}[W(s)]\leqslant\mathrm{E}[W(s-1)]-\mathrm{E}\left[\frac{\Delta^2(s)}{r(s)}\right]-\mathrm{E}[\xi(s)]+\mathrm{E}\left[\frac{\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2}{r^2(s)}(\varepsilon+\sigma^2)\right],\ \text{a.s.},$$

where

$$\xi(s):=\frac{\Phi^2(s-1)-1}{r(s)}\Delta^2(s)+\frac{r(s-1)\Phi^2(s-1)\Delta^2(s)}{r^2(s)}\left[1-2\frac{\phi(s-1)}{\Phi(s-1)}\Delta(s)\right]-\frac{\|\hat{\boldsymbol{\psi}}(s)\|^2}{r^2(s)}\Phi^2(s-1)\phi^2(s-1)\Delta^4(s).$$

Case I. If $\xi(s)\geqslant 0$, we have

$$\mathrm{E}[W(s)]\leqslant\mathrm{E}[W(s-1)]-\mathrm{E}\left[\frac{\Delta^2(s)}{r(s)}\right]+\mathrm{E}\left[\frac{\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2}{r^2(s)}(\varepsilon+\sigma^2)\right],\ \text{a.s.}\tag{3.22}$$

Case II. If $\xi(s)<0$ or $\zeta^2(s)>\varepsilon$, let $\hat{\boldsymbol{\theta}}(s):=\hat{\boldsymbol{\theta}}(s-1)$, then we have $\mathrm{E}[W(s)]=\mathrm{E}[W(s-1)]$, which ensures that the parameter estimation errors do not increase.

Note that

$$\sum_{s=1}^{\infty}\frac{\Phi^2(s-1)\|\hat{\boldsymbol{\psi}}(s)\|^2}{r^2(s)}\leqslant\sum_{s=1}^{\infty}\frac{r(s)-r(s-1)}{r(s-1)r(s)}=\frac{1}{r(0)}-\frac{1}{r(\infty)}<\infty.$$

Using the martingale convergence theorem [39] to (3.22), $\mathrm{E}[\|W(s)\|]$ converges to a constant C, and

$$\mathrm{E}\left[\sum_{s=1}^{\infty}\frac{\Delta^2(s)}{r(s)}\right]<\infty,\quad\text{or}\ \lim_{s\to\infty}\mathrm{E}\left[\frac{\Delta^2(s)}{r(s)}\right]=0.\tag{3.23}$$

Using (3.9) and (3.16) gives $\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+m)\tilde{\boldsymbol{\theta}}(s+m-1) = \Delta(s+m)$ and

$$\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+m)\tilde{\boldsymbol{\theta}}(s) = \Delta(s+m) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+m)\sum_{j=1}^{m-1}\frac{\hat{\boldsymbol{\psi}}(s+j)}{r(s+j)}\Phi(s+j)e(s+j).$$

Squaring and using (3.9) yields

$$\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(s)\hat{\boldsymbol{\psi}}(s+m)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+m)\tilde{\boldsymbol{\theta}}(s) \leqslant 2\Delta^2(s+m) + 4\|\hat{\boldsymbol{\psi}}(s+m)\|^2[\|\tilde{\boldsymbol{\theta}}(s+m-1)\|^2 + \|\tilde{\boldsymbol{\theta}}(s)\|^2].$$

Using (A2) gives

$$\begin{aligned}\alpha\|\tilde{\boldsymbol{\theta}}(s)\|^2 &\leqslant \tilde{\boldsymbol{\theta}}^{\mathrm{T}}(s)\left[\sum_{m=1}^{N-1}\frac{\hat{\boldsymbol{\psi}}(s+m)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+m)}{r(s+m)}\right]\tilde{\boldsymbol{\theta}}(s)\\ &\leqslant \sum_{j=1}^{N-1}\left[\frac{2\Delta^2(s+m)}{r(s+m)} + \frac{4\|\hat{\boldsymbol{\psi}}(s+m)\|^2}{r(s+m)}(\|\tilde{\boldsymbol{\theta}}(s+m-1)\|^2 + \|\tilde{\boldsymbol{\theta}}(s)\|^2)\right].\end{aligned}$$

Note that $\mathrm{E}[W(s)] = \mathrm{E}[\|\tilde{\boldsymbol{\theta}}(s)\|^2] \to C$, thus, we have

$$\lim_{s\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\theta}}(s)\|^2] \leqslant \frac{1}{\alpha}\lim_{s\to\infty}\mathrm{E}\left[\sum_{m=1}^{N-1}\frac{2\Delta^2(s+m)}{r(s+m)} + \frac{8C\|\hat{\boldsymbol{\psi}}(s+m)\|^2}{r(s+m)}\right]. \tag{3.24}$$

Using (4.20) gives the result of Theorem 1. □

Remark 2. The CRG method can be applied to the bilinear system with the right censored measurements, where the system output $y^*(s)$ is censored by

$$y(s) = \begin{cases} 0, & y^*(s) > 0\\ y^*(s). & y^*(s) \leqslant 0\end{cases}$$

Similar with the derivation of the left censored regression, when $y^*(s) \leqslant 0$, we have

$$\begin{aligned}\mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}, y(s) \leqslant 0] &= \boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} - \Omega\left(-\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}\right),\\ \mathrm{E}[y(s)|\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}] &= \Phi(-\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta} - \varphi(-\boldsymbol{\psi}^{\mathrm{T}}(s)\boldsymbol{\theta}).\end{aligned}$$

Under such a circumstance, we have

$$\hat{\boldsymbol{\theta}}(s) = \hat{\boldsymbol{\theta}}(s-1) + \frac{1}{r(s)}\Phi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\hat{\boldsymbol{\psi}}(s)e(s,\hat{\boldsymbol{\theta}}(s-1)), \tag{3.25}$$

$$r(s) = r(s-1) + \Phi^2(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\|\hat{\boldsymbol{\psi}}(s)\|^2, \tag{3.26}$$

$$e(s,\hat{\boldsymbol{\theta}}(s-1)) = y(s) - \Phi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1) + \varphi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)). \tag{3.27}$$

Equations (2.11)–(2.13) and (3.12)–(3.27) form the CRG method for the bilinear systems with right censored measurements. The performance of the CRG method in (2.11)–(2.13) and (3.12)–(3.27) can be analyzed by using the similar way as Theorem 1.

4. Censored regression-based multi-innovation gradient method

The CRG method can estimate the parameters and states of the bilinear state-space system with censored measurements. However, it may have a low estimation accuracy since only the current data $\{\boldsymbol{\psi}(s),\ y(s)\}$ and the innovation $e(s,\boldsymbol{\theta})$ are used at each instant. To enhance the estimation performance, the following derives the censored regression-based multi-innovation gradient method.

Define the cost function

$$J_2(\boldsymbol{\theta})=\frac{1}{2}\sum_{m=s-\rho+1}^{s}[y(m)-\mathrm{E}[y(m)|\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta}]]^2$$
$$=\frac{1}{2}\sum_{m=s-\rho+1}^{t}[y(m)-\Phi(\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta}-\varphi(\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta})]^2,$$

where ρ is the innovation length. We have

$$\frac{\partial J_2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=-\sum_{m=s-\rho+1}^{s}[\phi(\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta}+\Phi(\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta})-\phi(\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta})\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta}]\boldsymbol{\psi}(m)e(m,\boldsymbol{\theta})$$
$$=-\sum_{m=s-\rho+1}^{t}\Phi(\boldsymbol{\psi}^{\mathrm{T}}(m)\boldsymbol{\theta})\boldsymbol{\psi}(m)e(m,\boldsymbol{\theta}).$$

Define

$$\boldsymbol{Y}(\rho,s):=[y(s),y(s-1),\cdots,y(s-\rho+1)]^{\mathrm{T}}\in\mathbb{R}^{\rho},\tag{4.1}$$
$$\boldsymbol{E}(\rho,s):=[e(s,\hat{\boldsymbol{\theta}}(s-1)),e(s-1,\hat{\boldsymbol{\theta}}(s-1)),\cdots,e(s-\rho+1,\hat{\boldsymbol{\theta}}(s-1))]^{\mathrm{T}}\in\mathbb{R}^{\rho},$$
$$\boldsymbol{\Omega}(\rho,s):=[\hat{\boldsymbol{\psi}}(s),\hat{\boldsymbol{\psi}}(s-1),\cdots,\hat{\boldsymbol{\psi}}(s-\rho+1)]\in\mathbb{R}^{(M^2+2M)\times\rho},\tag{4.2}$$
$$\boldsymbol{\Gamma}(\rho,s):=[\varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)),\varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-1)\hat{\boldsymbol{\theta}}(s-1)),\cdots,\varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-\rho+1)\hat{\boldsymbol{\theta}}(s-1))]^{\mathrm{T}}\in\mathbb{R}^{\rho},\tag{4.3}$$
$$\boldsymbol{\Lambda}(\rho,s):=\mathrm{diag}\{\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)),\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-1)\hat{\boldsymbol{\theta}}(s-1)),\cdots,\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-\rho+1)\hat{\boldsymbol{\theta}}(s-1))\}\in\mathbb{R}^{\rho\times\rho}.\tag{4.4}$$

Using (3.6) gives

$$\boldsymbol{E}(\rho,s)=\boldsymbol{Y}(\rho,s)-\boldsymbol{\Lambda}(\rho,s)\boldsymbol{\Omega}^{\mathrm{T}}(\rho,s)\hat{\boldsymbol{\theta}}(s-1)-\boldsymbol{\Gamma}(\rho,s).\tag{4.5}$$

Minimizing $J_2(\boldsymbol{\theta})$ gives the calculation of the parameter vector $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}(s)=\hat{\boldsymbol{\theta}}(s-1)-\frac{1}{r(s)}\frac{\partial J_2(\hat{\boldsymbol{\theta}}(s-1))}{\partial\boldsymbol{\theta}}$$
$$=\hat{\boldsymbol{\theta}}(s-1)+\frac{1}{r(s)}\sum_{m=0}^{\rho-1}\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1))\hat{\boldsymbol{\psi}}(s-m)e(s-m,\hat{\boldsymbol{\theta}}(s-1))\tag{4.6}$$
$$=\hat{\boldsymbol{\theta}}(s-1)+\frac{1}{r(s)}\boldsymbol{\Omega}(\rho,s)\boldsymbol{\Lambda}(\rho,s)\boldsymbol{E}(\rho,s),\tag{4.7}$$
$$r(s)=r(s-1)+\max_{0\leqslant m\leqslant\rho-1}\{\Phi^2(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1))\|\hat{\boldsymbol{\psi}}(s-m)\|^2\},\tag{4.8}$$
$$\hat{\boldsymbol{\psi}}(s)=[-\hat{x}_M(s-M),-\hat{x}_{M-1}(s-M),\cdots,-\hat{x}_1(s-M),\hat{\boldsymbol{x}}^{\mathrm{T}}(s-1)u(s-1),$$
$$\hat{\boldsymbol{x}}^{\mathrm{T}}(s-2)u(s-2),\cdots,\hat{\boldsymbol{x}}^{\mathrm{T}}(s-M)u(s-M),u(s-1),u(s-2),\cdots,u(s-M)]^{\mathrm{T}},\tag{4.9}$$

$$\hat{\boldsymbol{\theta}}(s)=[\hat{g}_1(s),\hat{g}_2(s),\cdots,\hat{g}_M(s),\hat{\boldsymbol{f}}_1^{\mathrm{T}}(s),\hat{\boldsymbol{f}}_2^{\mathrm{T}}(s),\cdots,\hat{\boldsymbol{f}}_M^{\mathrm{T}}(s),\hat{h}_1(s),\hat{h}_2(s),\cdots,\hat{h}_M(s)]^{\mathrm{T}}. \tag{4.10}$$

Thus, Eqs (2.11)–(2.13) and (4.1)–(4.10) form CRMIG method for the bilinear state-space systems with left censored measurements in (2.1)–(2.3).

Theorem 2. For the CRMIG method in (2.11)–(2.13) and (4.1)–(4.10), suppose that $v(s)$ is a stochastic noise that follows the Gaussian, Uniform, or Laplace distribution and meets (A1), and there exist an integer N and a constant $\beta > 0$ irrelevant to s such that

$$\text{(A3)}\qquad \sum_{l=0}^{N-1}\sum_{m=0}^{\rho-1}\frac{\hat{\boldsymbol{\psi}}(s+l-m)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l-m)}{r(s+l)} \geqslant \beta\boldsymbol{I}. \tag{4.11}$$

Then, $\hat{\boldsymbol{\theta}}(s)$ can converge to its true value $\boldsymbol{\theta}$ in the mean-square sense.

Proof. Define $\Phi_m(s-1) := \Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1)) \in \mathbb{R}$, $\Phi_m(0) := \Phi(\boldsymbol{\psi}^{\mathrm{T}}(s-m)\boldsymbol{\theta}) \in \mathbb{R}$, $\varphi_m(s-1) := \varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1)) \in \mathbb{R}$, $\varphi_m(0) := \varphi(\boldsymbol{\psi}^{\mathrm{T}}(s-m)\boldsymbol{\theta}) \in \mathbb{R}$, $\tilde{\Phi}_m(s-1) := \Phi_m(s-1) - \Phi_m(0) \in \mathbb{R}$ and $\tilde{\varphi}_m(s-1) := \varphi_m(s-1) - \varphi_m(0) \in \mathbb{R}$. From (3.3), we have

$$\begin{aligned} y(s-m) &= \Phi_m(0)\boldsymbol{\psi}^{\mathrm{T}}(s-m)\boldsymbol{\theta} + \varphi_m(0) + v(s-m) \\ &= \Phi_m(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s-m)\boldsymbol{\theta} + \varphi_m(s-1) + \kappa(s-m), \end{aligned} \tag{4.12}$$

where

$$\kappa(s-m) := -\tilde{\Phi}_m(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s-m)\boldsymbol{\theta} - \tilde{\varphi}_m(s-1) + v(s-m) \in \mathbb{R}.$$

Using (4.6) and (4.12) gives

$$\begin{aligned} \tilde{\boldsymbol{\theta}}(s) &= \tilde{\boldsymbol{\theta}}(s-1) + \frac{1}{r(s)}\sum_{m=0}^{\rho-1}\Phi_m(s-1)\hat{\boldsymbol{\psi}}(s-m)[\Phi_m(s-1)\boldsymbol{\psi}^{\mathrm{T}}(s-m)\boldsymbol{\theta} + \kappa(s-m) \\ &\quad -\Phi_m(s-1)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1)], \\ &=: \tilde{\boldsymbol{\theta}}(s-1) + \frac{1}{r(s)}\sum_{m=0}^{\rho-1}\Phi_m(s-1)\hat{\boldsymbol{\psi}}(s-m) \\ &\quad \times[-\Phi_m(s-1)\Delta(s-m) + \Phi_m(s-1)\zeta(s-m) + \kappa(s-m)], \end{aligned} \tag{4.13}$$

where

$$\Delta(s-m) := \hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\tilde{\boldsymbol{\theta}}(s-1) \in \mathbb{R}, \tag{4.14}$$

$$\zeta(s-m) := [\boldsymbol{\psi}(s-m) - \hat{\boldsymbol{\psi}}(s-m)]^{\mathrm{T}}\boldsymbol{\theta} \in \mathbb{R}. \tag{4.15}$$

Using $(\sum_{m=1}^{M} a_m)^2 \leqslant M\sum_{m=1}^{M} a_m^2$ and taking the norm of both sides of (4.13) yields

$$\begin{aligned} \|\tilde{\boldsymbol{\theta}}(s)\|^2 &= \|\tilde{\boldsymbol{\theta}}(s-1) + \frac{1}{r(s)}\sum_{m=0}^{\rho-1}\Phi_m(s-1)\hat{\boldsymbol{\psi}}(s-m)[-\Phi_m(s-1)\Delta(s-m) \\ &\quad +\Phi_m(s-1)\zeta(s-m) + \kappa(s-m)]\|^2 \\ &= \|\tilde{\boldsymbol{\theta}}(s-1)\|^2 + \sum_{m=0}^{\rho-1}\frac{2}{r(s)}\Phi_m(s-1)\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(s-1)\hat{\boldsymbol{\psi}}(s-m)[-\Phi_m(s-1)\Delta(s-m) \end{aligned}$$

$$
\begin{aligned}
&+\Phi_m(s-1)\zeta(s-m)+\kappa(s-m)] \\
&+\left\|\frac{1}{r(s)}\sum_{m=0}^{\rho-1}\Phi_m(s-1)\hat{\boldsymbol{\psi}}(s-m)[-\Phi_m(s-1)\Delta(s-m)+\Phi_m(s-1)\zeta(s-m)+\kappa(s-m)]\right\|^2 \\
&\leqslant \|\tilde{\boldsymbol{\theta}}(s-1)\|^2+\sum_{m=0}^{\rho-1}\frac{2}{r(s)}\Phi_m(s-1)\Delta(s-m)[-\Phi_m(s-1)\Delta(s-m)+\Phi_m(s-1)\zeta(s-m) \\
&+\kappa(s-m)]+\rho\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}\{\Phi_m^2(s-1)\Delta^2(s-m) \\
&-2\Phi_m(s-1)\Delta(s-m)[\Phi_m(s-1)\zeta(s-m)+\kappa(s-m)]+[\Phi_m(s-1)\zeta(s-m)+\kappa(s-m)]^2\}.
\end{aligned}
$$

Define a nonnegative function $W(s) := \|\tilde{\boldsymbol{\theta}}(s)\|^2 \in \mathbb{R}$. Equation (4.8) means $r(s)-\rho\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2 \leqslant r(s)-\max_{0\leqslant m\leqslant\rho-1}\{\Phi_m^2(s-m)\|\hat{\boldsymbol{\psi}}(s-m)\|^2\} = r(s-1) \leqslant r(s)$ for a large ρ. Using (4.8) gives

$$
\begin{aligned}
&W(s) = W(s-1)-\sum_{m=0}^{\rho-1}\left[\frac{2}{r(s)}-\frac{\rho\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}\right]\Phi_m^2(s-1)\Delta^2(s-m) \\
&+2\sum_{m=0}^{\rho-1}\left[\frac{1}{r(s)}-\frac{\rho\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}\right]\Phi_m(s-1)\Delta(s-m)[\Phi_m(s-1)\zeta(s-m)+\kappa(s-m)] \\
&+\rho\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}[\Phi_m^2(s-1)\zeta^2(s-m)+2\Phi_m(s-1)\zeta(s-m)\kappa(s-m)+\kappa^2(s-m)] \\
&\leqslant W(s-1)-\sum_{m=0}^{\rho-1}\frac{1}{r(s)}\Phi_m^2(s-1)\Delta^2(s-m)-\sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}\Phi_m^2(s-1)\Delta^2(s-m) \\
&+2\sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}\Phi_m(s-1)\Delta(s-m)[\Phi_m(s-1)\zeta(s-m)+\kappa(s-m)] \\
&+\rho\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}[\Phi_m^2(s-1)\zeta^2(s-m)+2\Phi_m(s-1)\zeta(s-m)\kappa(s-m)+\kappa^2(s-m)] \\
&= W(s-1)-\sum_{m=0}^{\rho-1}\frac{1}{r(s)}\Phi_m^2(s-1)\Delta^2(s-m)-\sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}\Phi_m^2(s-1)[\Delta(s-m)-\zeta(s-m)]^2 \\
&+\sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}[\Phi_m^2(s-1)\zeta^2(s-m)+2\Phi_m(s-1)\Delta(s-m)\kappa(s-m)] \\
&+\rho\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}[\Phi_m^2(s-1)\zeta^2(s-m) \\
&+2\Phi_m(s-1)\zeta(s-m)\kappa(s-m)+\kappa^2(s-m)]. \qquad (4.16)
\end{aligned}
$$

Let $\phi_m(s-1) = \phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1))$. Then, $\kappa(s)$ can be expressed as

$$\kappa(s-m) = \phi_m(s-1)[\Delta(s-m)-\zeta(s-m)]^2+v(s-m). \tag{4.17}$$

Inserting (4.17) into (4.16) yields

$$\begin{aligned}
W(s) &= W(s-1) - \sum_{m=0}^{\rho-1}\frac{1}{r(s)}\Delta^2(s-m) - \sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)-1}{r(s)}\Delta^2(s-m)\\
&-\sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}\Phi_m^2(s-1)[\Delta(s-m)-\zeta(s-m)]^2\\
&+\sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}\{\Phi_m^2(s-1)\zeta^2(s-m)+2\Phi_m(s-1)\Delta(s-m)[\phi_m(s-1)(\Delta(s-m)-\zeta(s-m))^2\\
&+v(s-m)]\} + \rho\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}\{\Phi_m^2(s-1)\zeta^2(s-m)+2\Phi_m(s-1)\zeta(s-m)\\
&\times[\phi_m(s-1)(\Delta(s-m)-\zeta(s-m))^2+v(s-m)]+\phi_m^2(s-1)[\Delta(s-m)-\zeta(s-m)]^4\\
&+2\phi_m(s-1)[\Delta(s-m)-\zeta(s-m)]^2v(s-m)+v^2(s-m)\}.
\end{aligned} \tag{4.18}$$

Provided that $\zeta(s-m)$ is bounded with $\zeta^2(s-m) \leqslant \varepsilon$ for given $\varepsilon < \infty$, note that $r(s) \to \infty$, thus, $\zeta^2(s-m)/r(s) \to 0$. Since $\Delta(s)$, $\hat{\boldsymbol{\psi}}(s)$, $\zeta(s)$, and $r(s)$ are uncorrelated with $v(s)$, taking the expectation of both sides of (4.18) and using (A1), we have

$$\begin{aligned}
\mathrm{E}[W(s)] \leqslant \mathrm{E}[W(s-1)] - \mathrm{E}\left[\sum_{m=0}^{\rho-1}\frac{\Delta^2(s-m)}{r(s)}\right] - \mathrm{E}[\xi(s)]\\
+\rho\mathrm{E}\left[\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}(\varepsilon+\sigma^2)\right],\ \text{a.s.},
\end{aligned}$$

where

$$\begin{aligned}
\xi(s) := \sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)-1}{r(s)}\Delta^2(s-m) + \sum_{m=0}^{\rho-1}\frac{r(s-1)}{r^2(s)}\Phi_m^2(s-1)\Delta^2(s-m)\left[1-2\frac{\phi_m(s-1)}{\Phi_m(s-1)}\Delta(s-m)\right]\\
-\rho\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}\phi_m^2(s-1)\Delta^4(s-m).
\end{aligned}$$

Case I. If $\xi(s) \geqslant 0$, we have

$$\begin{aligned}
\mathrm{E}[W(s)] \leqslant \mathrm{E}[W(s-1)] - \mathrm{E}\left[\sum_{m=0}^{\rho-1}\frac{\Delta^2(s-m)}{r(s)}\right]\\
+\rho\mathrm{E}\left[\sum_{m=0}^{\rho-1}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}(\varepsilon+\sigma^2)\right],\ \text{a.s.}
\end{aligned} \tag{4.19}$$

Case II. If $\xi(s) < 0$ or $\zeta^2(s-m) > \varepsilon$, let $\hat{\boldsymbol{\theta}}(s) := \hat{\boldsymbol{\theta}}(s-1)$, then we have $\mathrm{E}[W(s)] = \mathrm{E}[W(s-1)]$, which ensures that the parameter estimation errors do not increase.

Note that

$$\sum_{s=1}^{\infty}\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)} \leqslant \sum_{s=1}^{\infty}\max_{0\leqslant m\leqslant \rho-1}\left\{\frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)}\right\}$$

$$= \sum_{s=1}^{\infty} \frac{r(s) - r(s-1)}{r(s-1)r(s)} = \frac{1}{r(0)} - \frac{1}{r(\infty)} < \infty.$$

Thus, we have

$$\sum_{s=1}^{\infty} \sum_{m=0}^{\rho-1} \frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)} = \sum_{m=0}^{\rho-1} \left[\sum_{s=1}^{\infty} \frac{\Phi_m^2(s-1)\|\hat{\boldsymbol{\psi}}(s-m)\|^2}{r^2(s)} \right] < \infty.$$

Using the martingale convergence theorem to (4.19), $\mathrm{E}[\|W(s)\|]$ converges to a constant C, and

$$\mathrm{E}\left[\sum_{m=0}^{\rho-1} \frac{\Delta^2(s-m)}{r(s)} \right] < \infty, \quad \text{or} \ \lim_{s\to\infty} \mathrm{E}\left[\sum_{m=0}^{\rho-1} \frac{\Delta^2(s-m)}{r(s)} \right] = 0. \tag{4.20}$$

Using (4.6) and (4.14) gives $\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l)\tilde{\boldsymbol{\theta}}(s+l-1) = \Delta(s+l)$ and

$$\begin{aligned} \hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l)\tilde{\boldsymbol{\theta}}(s) &= \Delta(s+l) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l) \sum_{k=1}^{j-1} \sum_{m=0}^{\rho-1} \frac{\Phi_m(s+k-1)}{r(s+k)} \hat{\boldsymbol{\psi}}(s-m+k)e(s-m+k) \\ &= \Delta(s+l) - \hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l)[\tilde{\boldsymbol{\theta}}(s+l-1) - \tilde{\boldsymbol{\theta}}(s)]. \end{aligned} \tag{4.21}$$

Squaring and using (3.9) yields

$$\tilde{\boldsymbol{\theta}}^{\mathrm{T}}(s)\hat{\boldsymbol{\psi}}(s+l)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l)\tilde{\boldsymbol{\theta}}(s) \leqslant 2\Delta^2(s+l) + 4\|\hat{\boldsymbol{\psi}}(s+l)\|^2[\|\tilde{\boldsymbol{\theta}}(s+l-1)\|^2 + \|\tilde{\boldsymbol{\theta}}(s)\|^2].$$

Using (A3) gives

$$\begin{aligned} \beta\|\tilde{\boldsymbol{\theta}}(s)\|^2 &\leqslant \tilde{\boldsymbol{\theta}}^{\mathrm{T}}(s) \left[\sum_{l=1}^{N-1} \sum_{m=0}^{\rho-1} \frac{\hat{\boldsymbol{\psi}}(s+l-m)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l-m)}{r(s+l)} \right] \tilde{\boldsymbol{\theta}}(s) \\ &= \sum_{l=1}^{N-1} \frac{1}{r(s+l)} \sum_{m=0}^{\rho-1} \tilde{\boldsymbol{\theta}}^{\mathrm{T}}(s)\hat{\boldsymbol{\psi}}(s+l-m)\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s+l-m)\tilde{\boldsymbol{\theta}}(s) \\ &\leqslant \sum_{l=1}^{N-1} \left[\sum_{m=0}^{\rho-1} \frac{2\Delta^2(s+l-m)}{r(s+l)} + \sum_{m=0}^{\rho-1} \frac{4\|\hat{\boldsymbol{\psi}}(s+l-m)\|^2}{r(s+l)} [\|\tilde{\boldsymbol{\theta}}(s+l-1)\|^2 + \|\tilde{\boldsymbol{\theta}}(s)\|^2] \right]. \end{aligned}$$

Note that $\mathrm{E}[W(s)] = \mathrm{E}[\|\tilde{\boldsymbol{\theta}}(s)\|^2] \to C$, thus, we have

$$\lim_{s\to\infty} \mathrm{E}[\|\tilde{\boldsymbol{\theta}}(s)\|^2] \leqslant \lim_{s\to\infty} \frac{1}{\beta} \sum_{l=1}^{N-1} \left[\sum_{m=0}^{\rho-1} \frac{2\Delta^2(s+l-m)}{r(s+l)} + \sum_{m=0}^{\rho-1} \frac{8C\|\hat{\boldsymbol{\psi}}(s+l-m)\|^2}{r(s+l)} \right].$$

Using (4.20) gives the result of Theorem 2. □

To ensure that the initial values have little impact on the parameter estimation, the unknown variables are usually specified to the small vectors close to zeros when the recursive algorithm begins. In this paper, $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{M^2+2M}/p_0 \in \mathbb{R}^{M^2+2M}$, $\hat{\boldsymbol{x}}(0) = \mathbf{1}_M/p_0$, where $p_0 = 10^6$.

Remark 3. In the CRMIG method in (2.11)–(2.13) and (4.1)–(4.10), the current measurement $\{\boldsymbol{\psi}(s),\ y(s)\}$ and the past $\rho-1$ measurements $\{\boldsymbol{\psi}(s-m),\ y(s-m),\ m = 1, 2, \cdots, \rho-1\}$ are used at each step. It means that the collected data are frequently used. Thus, the parameter and state estimation precision can be improved. When $\rho = 1$, the CRMIG method in (2.11)–(2.13) and (4.1)–(4.10) reduces to the CRG method in (2.11)–(2.13) and (3.9)–(3.13).

Remark 4. Similar with the derivation of the CRMIG method in (2.11)–(2.13) and (4.1)–(4.10) for the left censored measurements, the CRMIG method can be applied to the bilinear state-space systems with right censored measurements, where Eqs (4.3)–(4.5) and (4.8) should be modified as

$$\mathbf{\Gamma}(\rho, s) = [\varphi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)), \varphi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-1)\hat{\boldsymbol{\theta}}(s-1)), \cdots, \varphi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-\rho+1)\hat{\boldsymbol{\theta}}(s-1))], \tag{4.22}$$

$$\mathbf{\Lambda}(\rho, s) = \mathrm{diag}\{-\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1)), \Phi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-1)\hat{\boldsymbol{\theta}}(s-1)), \cdots, \Phi(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-\rho+1)\hat{\boldsymbol{\theta}}(s-1))\}, \tag{4.23}$$

$$\boldsymbol{E}(\rho, s) = \boldsymbol{Y}(\rho, s) - \mathbf{\Lambda}(\rho, s)\mathbf{\Omega}^{\mathrm{T}}(\rho, s)\hat{\boldsymbol{\theta}}(s-1) + \mathbf{\Gamma}(\rho, s), \tag{4.24}$$

$$r(s) = r(s-1) + \Phi^2(-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s)\hat{\boldsymbol{\theta}}(s-1))\|\hat{\boldsymbol{\psi}}(s)\|^2, \tag{4.25}$$

and other equations remain the same.

Table 2 demonstrates the computational complexities of the CRMIG method and the AM-LMS method by computing the total flops at each recursive step (take $v(s)$ as the Gaussian noise for an example here, and the other two cases are similar, i.e., $v(s)$ is uniform noise or Laplace noise).

Table 2. The computational complexities of the CRMIG method and the AM-RLS method.

Algorithms	Multiplications	Additions	Exponentiations	Total flops
AM-LMS	$5M^2 + 9M + 1$	$4M^2 + 8M - 1$	0	$N_1 := 9M^2 + 17M$
CRMIG	$3\rho(M^2 + 2M) + 4M^2$ $+7M + 10\rho + 2$	$3\rho(M^2 + 2M) + 2M^2$ $+4M - 1$	2	$N_2 := 6\rho(M^2 + 2M) + 6M^2$ $+11M + 10\rho + 3$

Note that

$$\begin{aligned} N_2 - N_1 &= 6\rho(M^2 + 2M) - 3M^2 - 6M + 10\rho + 3 \\ &= (6\rho - 3)(M^2 + 2M) + 10\rho + 3 > 0, \end{aligned}$$

thus, the CRMIG method requires more computational costs than the AM-LMS method. However, the CRMIG method has higher estimation precision than the AM-LMS method-see Figure 6.

The pseudo-code of implementing the CRMIG method for the left censored measurements is shown in Algorithm 1.

Algorithm 1 The CRMIG method.

Input: $\{u(s), y(s), s = 1, 2, \cdots, L\}$, M, ρ, ε.
Output: $\hat{\boldsymbol{\theta}}(s)$.

Initialization: $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{M^2+2M}/p_0$, $p_0 = 10^6$, $\hat{\boldsymbol{x}}(0) = \mathbf{1}_M/p_0$, $r(0) = 1$.
for $s = 1 : L$ **do**
    Form the information vector $\hat{\boldsymbol{\psi}}(s-m)$ $(m = 0, 1, 2, \cdots, \rho - 1)$ by (4.9).
    Compute the pdf $\varphi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1))$ and the cdf $\Phi(\hat{\boldsymbol{\psi}}^{\mathrm{T}}(s-m)\hat{\boldsymbol{\theta}}(s-1))$ by Table 1.
    Construct the information matrices/vectors $\boldsymbol{Y}(\rho, s)$, $\boldsymbol{\Omega}(\rho, s)$, $\boldsymbol{\Gamma}(\rho, s)$ and $\boldsymbol{\Lambda}(\rho, s)$ by (4.1)–(4.4).
    Compute $\boldsymbol{E}(\rho, s)$ and $r(s)$ by (4.5) and (4.8).
    Update $\hat{\boldsymbol{\theta}}(s)$ by (4.10).
    Read $\hat{g}_i(s)$, $\hat{\boldsymbol{f}}_i(s)$ and $\hat{h}_i(s)$ $(i = 1, 2, \cdots, n)$ from $\hat{\boldsymbol{\theta}}(s)$ by (4.10).
    Construct $\hat{\boldsymbol{G}}(s)$, $\hat{\boldsymbol{F}}(s)$ and $\hat{\boldsymbol{h}}(s)$ by (2.12) and (2.13).
    Compute $\hat{\boldsymbol{x}}(s)$ by (2.11).
    if $\|\hat{\boldsymbol{\theta}}(s) - \hat{\boldsymbol{\theta}}(s-1)\| > \varepsilon$
        $s := s + 1$;
    else
        Obtain $\hat{\boldsymbol{\theta}}(s)$, break;
    end
end for

5. Numerical simulation

Consider the following bilinear state-space system with censored measurements:

$$\begin{aligned}
\begin{bmatrix} x_1(s+1) \\ x_2(s+1) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -0.25 & 0.13 \end{bmatrix} \begin{bmatrix} x_1(s) \\ x_2(s) \end{bmatrix} + \begin{bmatrix} 0.10 & 0.12 \\ -0.15 & -0.05 \end{bmatrix} \begin{bmatrix} x_1(s) \\ x_2(s) \end{bmatrix} u(s) + \begin{bmatrix} 0.95 \\ -1.05 \end{bmatrix} u(s), \\
y^*(s) &= [1, 0] \begin{bmatrix} x_1(s) \\ x_2(s) \end{bmatrix} + v(s), \\
y(s) &= \begin{cases} y^*(s), & y^*(s) > 0, \\ 0, & y^*(s) \leqslant 0, \end{cases}
\end{aligned}$$

where the input $u(s)$ is taken as a persistent excitation signal with zero mean and unit variance. The true parameter vector to be estimated is

$$\begin{aligned}
\boldsymbol{\theta} &= [g_1, g_2, f_{11}, f_{12}, f_{21}, f_{22}, h_1, h_2]^{\mathrm{T}} \\
&= [-0.13, 0.25, 0.10, 0.12, -0.15, -0.05, 0.95, -1.05]^{\mathrm{T}}.
\end{aligned}$$

Take 3200 input-output data and $v(s)$ as the zero-mean Gaussian noise with variance $\sigma^2 = 0.50^2$. Choose the first $L = 3000$ input-output data for simulation and apply the CRMIG method to identify the bilinear system with censored measurements. It can be seen from Table 3 and Figure 1 that the CRMIG estimation error $\delta := \|\hat{\boldsymbol{\theta}}(s) - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\| \times 100\%$ decreases as s and ρ increase. It is evident that all the parameter estimates $\hat{g}_1(s)$, $\hat{g}_2(s)$, $\hat{f}_{11}(s)$, $\hat{f}_{12}(s)$, $\hat{f}_{21}(s)$, $\hat{f}_{22}(s)$, $\hat{h}_1(s)$, and $\hat{h}_2(s)$ can track their true values-see Figures 2 and 3.

Table 3. The CRMIG parameter estimation error with the different innovation length ρ.

ρ	s	g_1	g_2	f_{11}	f_{12}	f_{21}	f_{22}	h_1	h_2	δ (%)
1	100	-0.08567	0.08110	-0.03137	0.41315	0.07618	-0.06352	0.59758	-0.98821	38.30232
	200	-0.10410	0.08118	-0.00341	0.40167	0.05821	-0.06003	0.59878	-0.98312	36.89640
	500	-0.11688	0.07519	0.00989	0.40163	0.03389	-0.06339	0.61260	-0.98348	35.59981
	1000	-0.12720	0.08079	0.02770	0.39699	0.02834	-0.06184	0.62482	-0.98351	34.41298
	2000	-0.13770	0.08210	0.04231	0.39290	0.01858	-0.06345	0.63695	-0.98238	33.34309
	3000	-0.14043	0.08470	0.04792	0.38970	0.01201	-0.06023	0.64336	-0.98325	32.66814
2	100	-0.11379	0.11096	-0.03733	0.38829	0.03738	-0.03753	0.68656	-1.05419	31.75145
	200	-0.14745	0.11962	0.02858	0.35699	0.01104	-0.02986	0.68477	-1.02822	28.69745
	500	-0.16302	0.11865	0.05261	0.35025	-0.02614	-0.03155	0.70103	-1.01752	26.71873
	1000	-0.17450	0.13077	0.08164	0.33922	-0.03280	-0.02469	0.71709	-1.01282	25.11560
	2000	-0.18443	0.13576	0.09902	0.33149	-0.04680	-0.02549	0.73209	-1.00741	23.84163
	3000	-0.18447	0.14159	0.10340	0.32582	-0.05555	-0.02013	0.73939	-1.00697	23.02403
4	100	-0.11853	0.22683	0.06788	0.13087	-0.10614	0.01968	0.91437	-1.09079	7.35123
	200	-0.15261	0.24844	0.13440	0.11204	-0.13827	0.02232	0.90930	-1.05448	6.42367
	500	-0.15437	0.23708	0.10839	0.14103	-0.13981	-0.00813	0.91847	-1.05776	4.42837
	1000	-0.15465	0.24455	0.11709	0.13635	-0.13807	-0.00826	0.92288	-1.05402	4.23822
	2000	-0.15445	0.23675	0.10720	0.13734	-0.14370	-0.02019	0.92574	-1.05016	3.51917
	3000	-0.14986	0.23550	0.09895	0.13635	-0.14426	-0.02101	0.92319	-1.05074	3.39962
6	100	-0.11371	0.26736	0.06910	0.10557	-0.13610	0.02621	0.96054	-1.06899	6.20007
	200	-0.14967	0.27314	0.13355	0.09054	-0.16444	0.02174	0.93843	-1.03518	6.35760
	500	-0.14815	0.24785	0.09394	0.12787	-0.15336	-0.02412	0.94111	-1.04668	2.37502
	1000	-0.14668	0.25546	0.10852	0.12202	-0.14610	-0.02150	0.94372	-1.04563	2.44043
	2000	-0.14591	0.24183	0.09675	0.12337	-0.15139	-0.03691	0.94473	-1.04277	1.67005
	3000	-0.13969	0.23983	0.08731	0.12181	-0.15059	-0.03696	0.93900	-1.04454	1.78909
True values		-0.13000	0.25000	0.10000	0.12000	-0.15000	-0.05000	0.95000	-1.05000	

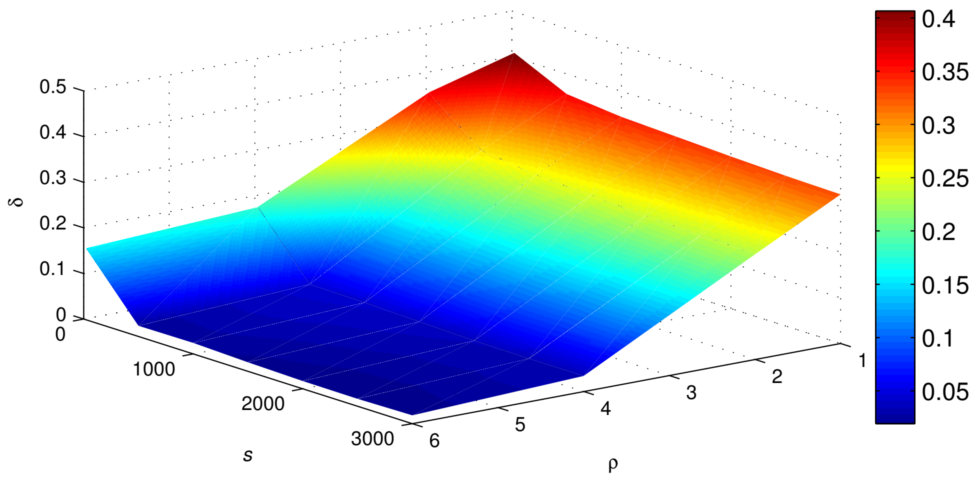

Figure 1. The parameter estimation error of the CRMIG method.

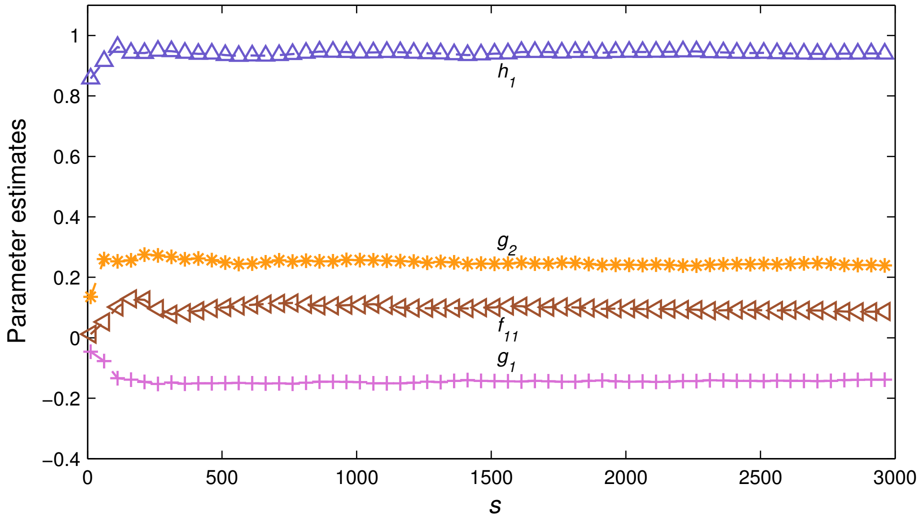

Figure 2. The parameter estimates $\hat{g}_1(s)$, $\hat{g}_2(s)$, $\hat{f}_{11}(s)$, and $\hat{h}_1(s)$ versus s.

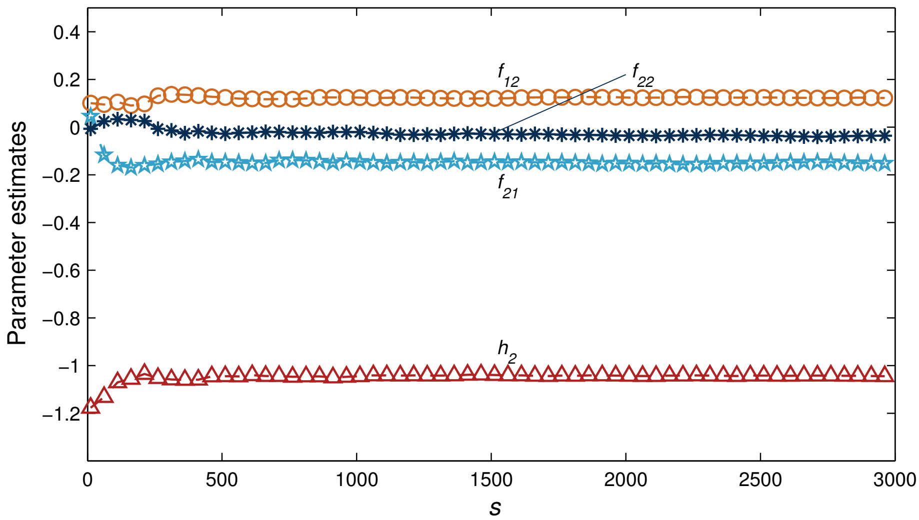

Figure 3. The parameter estimates $\hat{f}_{12}(s)$, $\hat{f}_{21}(s)$, $\hat{f}_{22}(s)$, and $\hat{h}_2(s)$ versus s.

To test the effect of noise variance variation on the CRMIG method, Table 4 and Figure 4 are given. As can be observed, as the noise variance becomes small, the CRMIG method has lower parameter estimation errors.

Table 4. The CRMIG parameter estimation error under the Gaussian noise variance $\sigma^2 = 1.00^2$.

s	g_1	g_2	f_{11}	f_{12}	f_{21}	f_{22}	h_1	h_2	δ (%)
100	-0.04992	0.21200	0.05199	0.13661	-0.08239	0.00397	0.97774	-1.09259	9.80315
200	-0.10918	0.25713	0.19817	0.09170	-0.15896	0.02421	0.95235	-1.01272	9.16027
500	-0.10944	0.23044	0.12052	0.16282	-0.15843	-0.04719	0.95946	-1.02958	4.13327
1000	-0.11058	0.24219	0.14619	0.14798	-0.13829	-0.04278	0.96022	-1.02902	4.37713
2000	-0.11304	0.21891	0.12971	0.14904	-0.14871	-0.06769	0.96177	-1.02679	4.31451
3000	-0.10308	0.21354	0.11007	0.14266	-0.14986	-0.06350	0.94727	-1.03297	3.84130
	-0.13000	0.25000	0.10000	0.12000	-0.15000	-0.05000	0.95000	-1.05000	

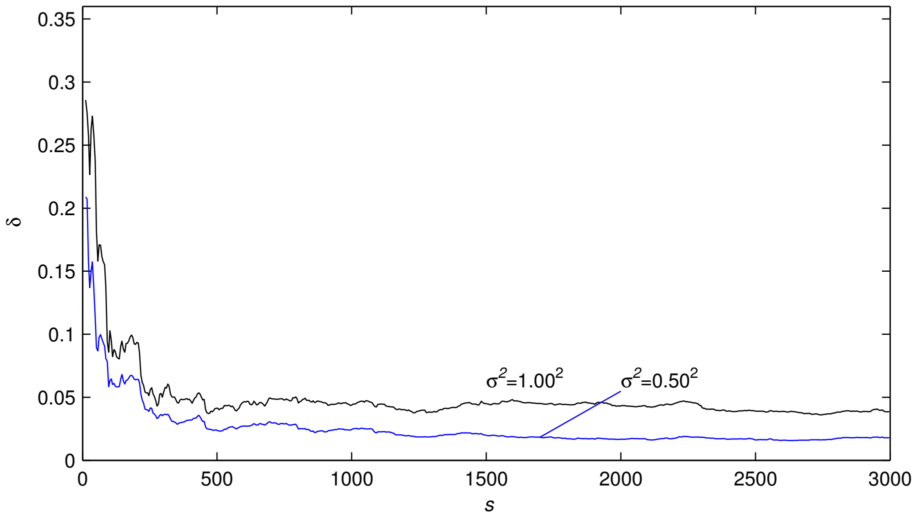

Figure 4. The CRMIG parameter estimation error under different Gaussian noise variances.

To comprehensively compare the differences of the CRMIG method under different settings, apart from the aforementioned parameter estimation error index $\delta = \|\hat{\boldsymbol{\theta}}(s) - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\| \times 100\%$, some other indexes are introduced, such as the state estimation error index ϵ and the predicted output root-mean-square error (RMSE) index τ, which are defined as

$$\epsilon := \left[\frac{1}{L}\sum_{m=1}^{L}\frac{\|\hat{\boldsymbol{x}}(m) - \boldsymbol{x}(m)\|^2}{\|\boldsymbol{x}(m)\|^2}\right]^{\frac{1}{2}} \times 100\%,$$

$$\tau := \left[\frac{1}{L_r}\sum_{m=L+1}^{L+L_r}[\hat{y}(m) - y(m)]^2\right]^{\frac{1}{2}},$$

where $L_r = 200$ represents the the rest data from 3001 to 3200 used for model validation. Table 5 exhabits the comparisons of these three indexes under different innovation lengths and noise variances. As can be observed, three indexes decrease with ρ increasing and σ^2 decreasing, which indicates that a large ρ and a small σ^2 can be beneficial to the improvements of the parameter and state estimation precision and the prediction accuracy.

Table 5. The comparisons of some indexes under different settings.

Indexes	$\rho = 5,\ \sigma^2 = 0.50^2$	$\rho = 6,\ \sigma^2 = 0.50^2$	$\rho = 6,\ \sigma^2 = 1.00^2$
δ (%)	2.25785	1.78909	3.84130
ϵ (%)	0.37202	0.33273	0.57168
τ	1.05836	1.05732	1.18309

To test the performance of the CRMIG method under different background noise environments, the stochastic noise $v(s)$ is also chosen as the uniformly distributed noise over the different internal $[-a, a]$ and the Laplace noise with the pdf $\phi(\tau) = \frac{1}{2\lambda}\mathrm{e}^{-\frac{|\tau|}{\lambda}}$ and $\lambda = 1$, respectively. Under the innovation length $\rho = 6$, Table 6 and Figure 5 show the corresponding parameter estimation errors. As shown in Figure 5, the CRMIG method achieves good estimation performance under different kinds of background noises.

Table 6. The CRMIG estimation error under the different noise distributions ($\rho = 6$).

$v(s)$	s	g_1	g_2	f_{11}	f_{12}	f_{21}	f_{22}	h_1	h_2	δ (%)
Uniform noise	100	-0.14669	0.15963	0.04294	0.10831	-0.05293	-0.07065	0.96312	-0.91243	13.82507
	200	-0.13585	0.14824	0.06045	0.12345	-0.13069	-0.07075	0.95392	-0.98289	9.00103
	500	-0.16134	0.20151	0.07981	0.10080	-0.12491	-0.08155	0.97532	-0.98196	7.18189
	1000	-0.14072	0.24035	0.07534	0.10464	-0.13121	-0.06489	0.96450	-1.00870	4.07472
	2000	-0.13392	0.23039	0.08776	0.10095	-0.13913	-0.05990	0.96764	-1.01547	3.51214
	3000	-0.14015	0.23537	0.09192	0.10828	-0.13943	-0.05868	0.96709	-1.02245	2.86994
Laplace noise	100	-0.03719	-0.13847	0.10839	-0.17329	-0.00705	-0.37801	0.68859	-1.12456	45.79703
	200	-0.05211	-0.01639	0.05653	-0.00266	-0.02631	-0.20381	0.79777	-1.17316	28.33057
	500	-0.11692	0.11625	0.01263	0.08387	-0.06343	-0.14061	0.88404	-1.14178	16.12489
	1000	-0.14547	0.18330	0.08078	0.08594	-0.10487	-0.09709	0.92987	-1.09813	7.86315
	2000	-0.14580	0.24659	0.08722	0.09393	-0.11692	-0.06881	0.95149	-1.05381	3.46985
	3000	-0.14240	0.23738	0.08016	0.09797	-0.11588	-0.04650	0.95040	-1.03695	3.44979
True values		-0.13000	0.25000	0.10000	0.12000	-0.15000	-0.05000	0.95000	-1.05000	

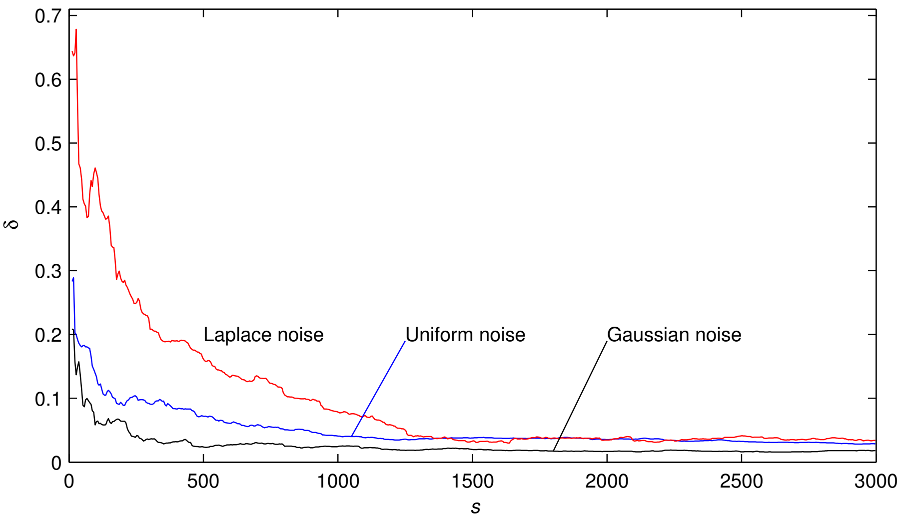

Figure 5. The CRMIG parameter estimation error under the different noise distributions.

To show the advantage of the CRMIG method, Table 7 and Figure 6 compare the CRMIG method ($\rho = 6$) with the AM-LMS method under the same conditions. It can be obviously observed from Figure 6 that the CRMIG method performs better than the AM-LMS method. Figures 7 and 8 depict the true states, the predicted AM-LMS state estimates, and the predicted CRMIG state estimates. As can be seen, the predicted CRMIG state estimates are closer the true states than the AM-LMS state estimates. It can be inferred that the CRMIG method is superior than the AM-LMS method for estimating the parameters and states of the bilinear system with censored measurements.

Table 7. The parameter estimation error by the AM-LMS method.

s	g_1	g_2	f_{11}	f_{12}	f_{21}	f_{22}	h_1	h_2	δ (%)
100	-0.11458	0.26127	0.02858	0.24034	-0.07485	0.02226	0.71008	-0.65540	33.82164
200	-0.12900	0.25731	0.06220	0.24392	-0.12586	0.02386	0.69307	-0.65225	34.02923
500	-0.13391	0.24783	0.05540	0.24005	-0.11974	-0.03443	0.68872	-0.67530	32.56210
1000	-0.13042	0.25159	0.08001	0.24475	-0.10428	-0.03363	0.68496	-0.68589	32.19328
2000	-0.13477	0.23848	0.08140	0.24874	-0.10326	-0.04489	0.67794	-0.69597	32.01043
3000	-0.13443	0.23192	0.07922	0.24790	-0.10839	-0.04573	0.67148	-0.70201	31.93081
	-0.13000	0.25000	0.10000	0.12000	-0.15000	-0.05000	0.95000	-1.05000	

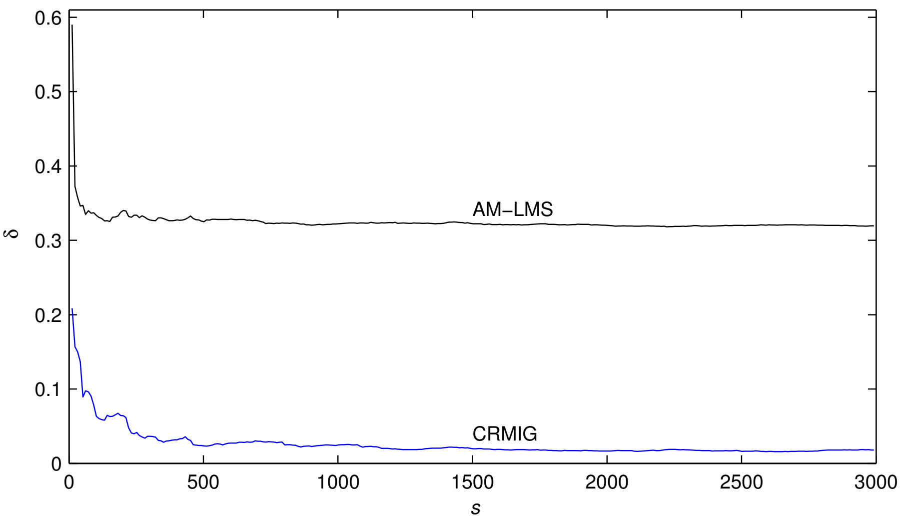

Figure 6. The parameter estimation errors of the CRMIG method and the AM-LMS method.

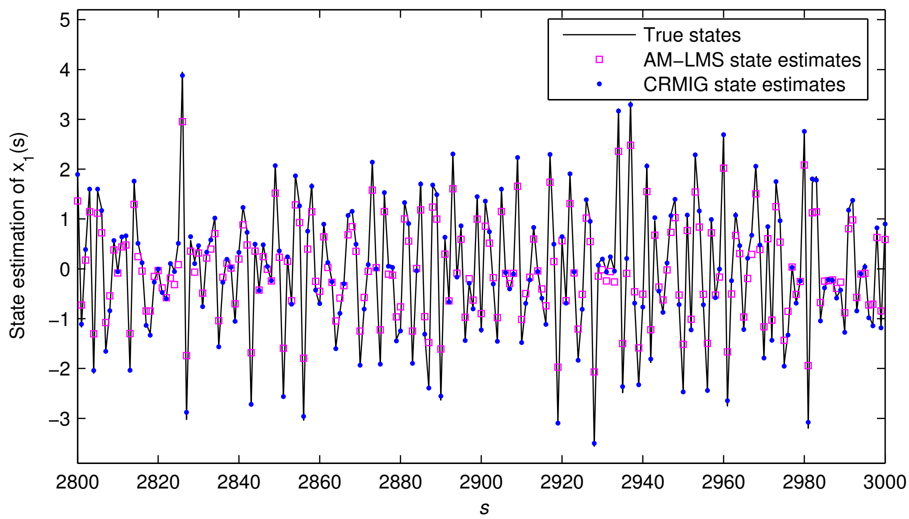

Figure 7. The true states $x_1(s)$ and the estimated states $\hat{x}_1(s)$ given by the AM-LMS method and the CRMIG method.

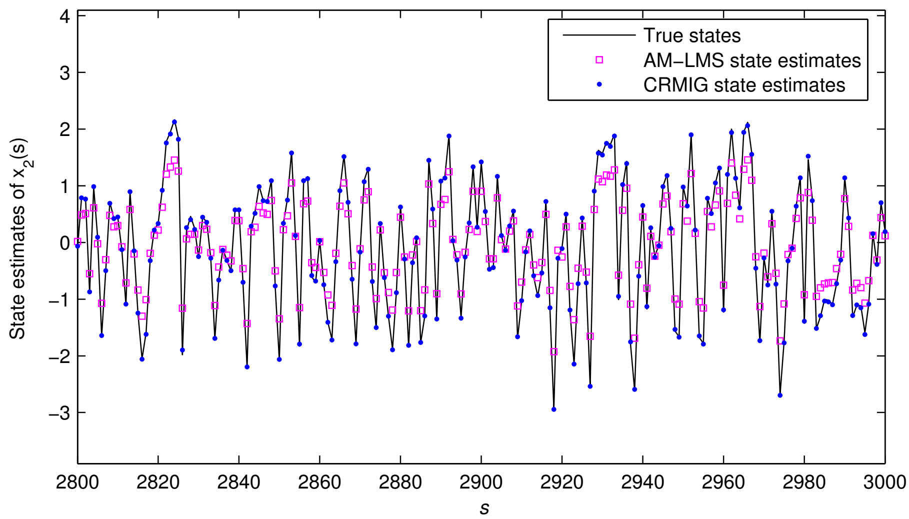

Figure 8. The true states $x_2(s)$ and the estimated states $\hat{x}_2(s)$ given by the AM-LMS method and the CRMIG method.

Figure 9 compares the true outputs, the censored outputs, the predicted AM-LMS outputs and the predicted CRMIG outputs. As can be seen from Figure 9, the predicted CRMIG outputs are nearer to the true outputs than the predicted AM-LMS outputs, which demonstrates that the CRMIG method has better tracking performance than the AM-LMS method for the bilinear system with censored measurements, since the CRMIG method utilizes more prior statistical knowledge about the measurement noise $v(s)$.

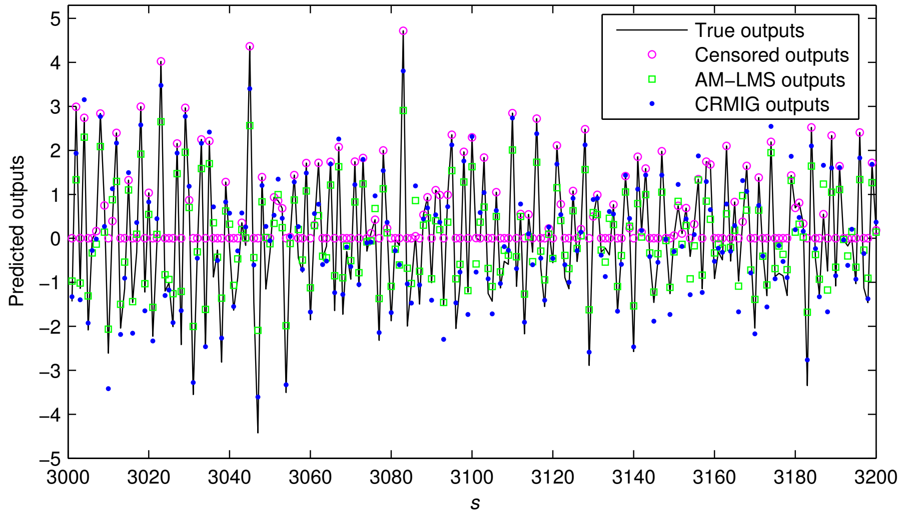

Figure 9. The true outputs, the censored outputs, the predicted AM-LMS outputs, and the predicted CRMIG outputs.

6. Conclusions

In this paper, a censored regression-based gradient method is proposed for estimating the parameters and states of the bilinear state-space systems with censored measurements based on the probit regression. The method adds the bias compensation term into the recursive expression for avoiding the biased parameter estimation. To enhance the performance, a censored regression-based multi-innovation gradient method is developed, where the repeated use of the input-output data and innovations guarantees the improvement of the estimation accuracy. Simulation results further verify the superiority of the proposed methods. In the future, the basic idea of the proposed method can be combined with other approaches such as the particle filter and the expectation maximization method to study the identification issues of other nonlinear state-space systems with censored measurements.

Author contributions

Xuehai Wang: Methodology, writing-review & editing, writing-original draft; Fang Zhu: Validation, writing-review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Science and Technology Project of Henan Province, China (No. 202102210297).

Conflict of interest

The authors declare no potential conflict of interests.

References

1. N. T. Jamba, G. Jacinto, P. A. Filipe, C. A. Braumann, Estimation for stochastic differential equation mixed models using approximation methods, *AIMS Math.*, **9** (2024), 7866–7894. https://dx.doi.org/10.3934/math.2024383
2. F. Ding, L. Xu, X. Zhang, H. Xu, Y. H. Zhou, X. L. Luan, Hierarchical generalized extended parameter identification for multivariable equation-error ARMA-like systems by using the filtering identification idea, *Annu. Rev. Control*, **60** (2025), 100993. https://doi.org/10.1016/j.arcontrol.2025.100993
3. P. Zhong, X. L. Wu, L. Zhu, A. H. Yang, A new APSO-SPC method for parameter identification problem with uncertainty caused by random measurement errors, *AIMS Math.*, **10** (2025), 3848–3865. https://doi.org/10.3934/math.2025179
4. G. Q. Miao, F. Ding, Q. Y. Liu, E. F. Yang, Iterative parameter identification algorithms for transformed dynamic rational fraction input-output systems, *J. Comput. Appl. Math.*, **434** (2023), 115297. https://doi.org/10.1016/j.cam.2023.115297
5. T. C. Zong, J. H. Li, G. P. Lu, Parameter identification of dual-rate Hammerstein-Volterra nonlinear systems by the hybrid particle swarm-gradient algorithm based on the auxiliary model, *Eng. Appl. Artif. Intel.*, **117** (2023), 105526. https://doi.org/10.1016/j.engappai.2022.105526
6. A. Carini, R. Forti, S. Orcioni, A polynomial multiple variance method for Volterra filter identification, *IEEE T. Instrum. Meas.*, **74** (2025), 1–12. https://doi.org/10.1109/TIM.2025.3546397
7. X. H. Wang, Y. J. Duan, Heterogeneous mixed kernel correntropy-based robust iterative estimation methods and convergence analysis for the nonlinear system with outliers, *Circuits Syst. Signal Process.*, 2025. https://doi.org/10.1007/s00034-025-03156-z
8. Y. Gu, J. J. Chen, Y. H. Tan, A scheme of gradient-based iterative identification with separable synchronous variable-innovation for Exp-ARX systems with arbitrary data loss, *IEEE T. Ind. Inform.*, 2025. https://doi.org/10.1109/TII.2025.3574425
9. T. Gonzàlez, A. Blais, N. Couëllan, C. Ruiz, Distributional loss for convolutional neural network regression and application to parameter estimation in satellite navigation signals, *Expert Syst. Appl.*, **255** (2024), 124682. https://doi.org/10.1016/j.eswa.2024.124682
10. X. H. Wang, Y. G. Liu, S. R. Zhao, Robust fixed-point Kalman smoother for bilinear state-space systems with non-Gaussian noise and parametric uncertainties, *Int. J. Adapt. Control Signal Process.*, **38** (2024), 3636–3655. https://doi.org/10.1002/acs.3891
11. R. Larhrissi, M. Benoudi, Regional stabilization of a fractional output for a class of time-delayed distributed bilinear systems, *Int. J. Dynam. Control*, **12** (2024), 992–1002. https://doi.org/10.1007/s40435-023-01232-3

12. I. El Haskouki, A. Tsouli, M. Ouarit, Uniform exponential stabilization for distributed discrete bilinear systems with time delay, *J. Differ. Equ. Appl.*, **31** (2025), 549–569. https://doi.org/10.1080/10236198.2024.2446457

13. G. Y. Cheng, Z. Miao, Y. L. Jiang, High-order Krylov subspace model order reduction methods for bilinear time-delay systems, *Syst. Control Lett.*, **186** (2024), 105764. https://doi.org/10.1016/j.sysconle.2024.105764

14. A. Ghezal, O. Alzeley, Probabilistic properties and estimation methods for periodic threshold autoregressive stochastic volatility, *AIMS Math.*, **9** (2024), 11805–11832. https://doi.org/10.3934/math.2024578

15. F. Ding, X. L. Luan, L. Xu, X. Zhang, Hierarchical recursive gradient parameter identification for multi-input ARX systems with partially-coupled information vectors, *Int. J. Adapt. Control Signal Process.*, **39** (2025). https://doi.org/10.1002/acs.4036

16. L. Xu, X. Zhang, F. Ding, Q. M. Zhu, A delta operator state estimation algorithm for discrete-time systems with state time-delay, *IEEE Signal Process. Lett.*, **32** (2025), 391–395. https://doi.org/10.1109/LSP.2024.3519897

17. S. An, L. J. Wang, Y. He, Filtering-based maximum likelihood hierarchical recursive identification algorithms for bilinear stochastic systems, *Nonlinear Dyn.*, **111** (2023), 12405–12420. https://doi.org/10.1007/s11071-023-08458-y

18. M. H. Li, X. M. Liu, Y. M. Fan, F. Ding, Highly efficient maximum-likelihood identification methods for bilinear systems with colored noises, *J. Syst. Control Eng.*, **238** (2024), 1763–1784. https://doi.org/10.1177/09596518241256145

19. Y. Gu, W. Dai, Q. M. Zhu, H. Nouri, Hierarchical multi-innovation stochastic gradient identification algorithm for estimating a bilinear state-space model with moving average noise, *J. Comput. Appl. Math.*, **420** (2023), 114794. https://doi.org/10.1016/j.cam.2022.114794

20. W. J. Wang, S. Y. Liu, Y. H. Jiang, J. F. Sun, W. X. Xu, X. H. Chen, et al., A robust filter and smoother-based expectation-maximization algorithm for bilinear systems with heavy-tailed noise, *Mech. Syst. Signal Process.*, **236** (2025), 112912. https://doi.org/10.1016/j.ymssp.2025.112912

21. S. Y. Liu, Y. J. Wang, F. Ding, A. Alsaedi, T. Hayat, Joint iterative state and parameter estimation for bilinear systems with autoregressive noises via the data filtering, *ISA Trans.*, **147** (2024), 337–349. https://doi.org/10.1016/j.isatra.2024.01.035

22. W. H. Song, Z. D. Wang, Z. K. Li, H. L. Dong, Q. L. Han, Protocol-based particle filtering for nonlinear complex networks: Handling non-Gaussian noises and measurement censoring, *IEEE T. Netw. Sci. Eng.*, **10** (2023), 128–139. https://doi.org/10.1109/TNSE.2022.3205553

23. L. Zhao, G. Y. Chen, X. Y. Cao, L. Li, Distributed fusion for multisensor systems with censored measurements under Round-Robin protocol, *IEEE Sens. J.*, **25** (2025), 19888–19895. https://doi.org/10.1109/JSEN.2025.3560117

24. H. P. Ren, Z. W. Zhang, Q. Gong, Estimation of Shannon entropy of the inverse exponential Rayleigh model under progressively Type-II censored test, *AIMS Math.*, **10** (2025), 9378–9414. https://doi.org/10.3934/math.2025434

25. J. X. Li, J. Hu, H. J. Liu, H. Yu, Z. H. Wu, Nonlinear distributed filtering subject to censoring measurements under dynamic event-triggered communication mechanism: The state-saturated case, *Commun. Nonlinear Sci. Numer. Simulat.*, **114** (2022), 106618. https://doi.org/10.1016/j.cnsns.2022.106618

26. J. Cheng, Q. W. Zhang, D. Zhang, H. C. Yan, W. H. Qi, Asynchronous filtering for fuzzy sampled-data systems subject to censored measurements and non-uniform sampling, *IEEE T. Fuzzy Syst.*, **32** (2024), 5296–5306. https://doi.org/10.1109/TFUZZ.2024.3420793

27. T. I. Lin, W. L. Wang, Finite mixtures of multivariate contaminated normal censored regression models, *J. Comput. Graph. Stat.*, 2025, 1–14. https://doi.org/10.1080/10618600.2025.2495257

28. H. Q. Zhao, G. Wang, F. Zhao, D. X. Liu, P. C. Song, Recursive general mixed norm algorithm for censored regression: Performance analysis and channel equalization application, *IEEE T. Syst. Man Cybern.: Syst.*, **54** (2024), 752–763. https://doi.org/10.1109/TSMC.2023.3321115

29. G. B. Qian, L. P. Shen, Y. H. Guan, J. H. Qian, S. Y. Wang, Robust recursive widely linear adaptive filtering algorithm for censored regression, *Signal Process.*, **230** (2025), 109854. https://doi.org/10.1016/j.sigpro.2024.109854

30. F. Ding, L. Xu, X. Zhang, Y. H. Zhou, X. L. Luan, Recursive identification methods for general stochastic systems with colored noises by using the hierarchical identification principle and the filtering identification idea, *Annu. Rev. Control*, **57** (2024), 100942. https://doi.org/10.1016/j.arcontrol.2024.100942

31. H. H. Wang, X. M. Liu, Highly efficient three-stage maximum likelihood recursive least squares identification method for multiple-input multiple-output systems, *Syst. Control Lett.*, **200** (2025), 106094. https://doi.org/10.1016/j.sysconle.2025.106094

32. F. Ding, L. Xu, P. Liu, X. F. Wang, Two-stage parameter estimation methods for linear time-invariant continuous-time systems, *Syst. Control Lett.*, **204** (2025), 106166. https://doi.org/10.1016/j.sysconle.2025.106166

33. G. Mzyk, A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems, *IEEE T. Circuits Syst. II Express Briefs*, **54** (2007), 897–901. https://doi.org/10.1109/TCSII.2007.901634

34. E. M. Almetwally, H. M. Almongy, A. E. sayed Mubarak, Bayesian and maximum likelihood estimation for the Weibull generalized exponential distribution parameters using progressive censoring schemes, *Pak. J. Stat. Oper. Res.*, **14** (2018), 853–868. https://doi.org/10.18187/pjsor.v14i4.2600

35. Z. Y. Wang, Z. L. Liu, Y. T. Dai, K. J. Cai, Recursive estimation for system with random transmission loss and censored measurement, *IEEE T. Circuits Syst. II Express Briefs*, **70** (2022), 2281–2285. https://doi.org/10.1109/TCSII.2022.3226485

36. S. Kim, H. Kim, S. Lee, Adaptive censoring-probability-based Tobit Kalman filter for tracking maneuvering targets, *Aerosp. Sci. Technol.*, **163** (2025), 110263. https://doi.org/10.1016/j.ast.2025.110263

37. G. B. Qian, L. P. Shen, Y. H. Guan, J. H. Qian, S. Y. Wang, Robust recursive widely linear adaptive filtering algorithm for censored regression, *Signal Process.*, **230** (2025), 109854. https://doi.org/10.1016/j.sigpro.2024.109854
38. G. Wang, H. Q. Zhao, Robust adaptive least mean M-estimate algorithm for censored regression, *IEEE T. Syst. Man Cybern.: Syst.*, **52** (2022), 5165–5174. https://doi.org/10.1109/TSMC.2021.3120272
39. F. Ding, Y. Gu, Performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state-space systems with one-step state delay, *Circuits Syst. Signal Process.*, **32** (2013), 585–599. https://doi.org/10.1007/s00034-012-9463-5

AIMS Press