SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

January 2023, Vol. 66 112204:1–112204:20
https://doi.org/10.1007/s11432-021-3505-5

System identification under saturated precise or set-valued measurements

Yanlong ZHAO[1,2*], Hang ZHANG[1,2], Ting WANG[3] & Guolian KANG[4]

[1]*Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China;*
[2]*School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, China;*
[3]*Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing* 100083*, China;*
[4]*Department of Biostatistics, St. Jude Children's Research Hospital, Memphis* TN 38105*, USA*

Received 27 September 2021/Revised 26 January 2022/Accepted 22 April 2022/Published online 28 December 2022

Abstract This paper considers the system identification problem based on saturated precise or set-valued measurements, which is widely used in various fields and has essential difficulties. Since existing methodologies cannot make full use of the mixed data, this paper is aiming to fill the gap and build a unified framework in dealing with such problems rigorously and comprehensively. New algorithms are introduced and their properties are established. Most significantly, the Cramér-Rao (CR) lower bound based on the measurements is established, which consists of two parts with respect to the precise data and set-valued data, respectively. This prompts the idea of designing an estimation algorithm by grouping and combining the estimations under two classifications of data. As a result, a CR lower bound-based algorithm (CRBA) is constructed. The convergence properties are theoretically analyzed in terms of consistency and asymptotic efficiency under periodic inputs. For general inputs, an algorithm based on the CRBA that combines the expectation maximization (EM) algorithm for set-valued subsystems and the gradient descent algorithm for precise subsystems is proposed. Numerical simulations validate the superiority of the proposed algorithms.

Keywords system identification, Cramér-Rao lower bound, truncated data, precise measurement, set-valued measurement

Citation Zhao Y L, Zhang H, Wang T, et al. System identification under saturated precise or set-valued measurements. Sci China Inf Sci, 2023, 66(1): 112204, https://doi.org/10.1007/s11432-021-3505-5

1 Introduction

System identification aims to construct the mathematical models of dynamic systems from the measurements of system inputs and outputs [1], which has been widely applied to industrial processes, control systems, medicine, and many other fields. As one of the elements of system identification [2], data play an important role. Depending on the different measurements, system identification problems have been formulated for precise data as well as set-valued data, which supply information that whether the system outputs belong to some certain sets [3]. Accordingly, some fundamental frameworks and methods have been developed.

This paper considers system identification problems under saturated precise or set-valued measurements, which are motivated with wide applications. At each time, the sensor might show different types of measurements: a precise value if the system output with stochastic noise lies in some certain work zones of the sensor, and otherwise set-valued one. To help make sense of the characters of the measurements, specific examples for both classical systems and emerging technologies will be introduced in the following subsection.

The mixture of precise and set-valued data introduces highly non-trivial technical challenges. (a) How should the data be grouped and mixed? (b) How should we develop weights on data to reduce identification errors? (c) Can we develop optimal algorithms under the mixed data? To solve these system

* Corresponding author (email: ylzhao@amss.ac.cn)

identification problems in this case, we may apply classic methods by only using precise data, or set-valued methods by considering all data as set-valued ones. Both methods only utilize one type of the data and cannot make full use of the measurements.

This paper is aiming to fill the gap between the mixed data and existing methodologies, and as far as we know, is the first comprehensive framework in dealing with such problems rigorously and comprehensively. New algorithms are developed and their properties are established.

1.1 Motivations and background

Firstly, we introduce some examples of saturated precise or set-valued measurements, which widely emerge in practical fields and motivate this research.

(i) Hall-effect sensors [4–6]. In engineering science, sensor saturation arises frequently in measurements as a result of physical limitations on the sensors. A typical example is the Hall-effect sensor, based on the famous "Hall effect" principle, that is widely used in products, ranging from computers to sewing machines, from automobiles to aircraft, and from machine tools to medical equipment. The sensor responds with an output voltage proportional to the magnetic field input. An additional amplifier serves as a core unit between the input and output, and the output of the amplifier cannot exceed the limits imposed by the power supply. Common Hall-effect sensors can measure precise voltages in the ranges of 4.5 to 10.5, 4.5 to 12, or 6.6 to 12.6 DCV (direct current voltage), respectively. Otherwise, the sensor can supply only set-valued information, e.g., when the voltage is less than 4.5 or larger than 10.5. The saturation occurs in the amplifier and not in the Hall element; thus, large magnetic fields will not damage the Hall-effect sensors but rather drive them into saturation.

(ii) Biosystems with survival data [7, 8] or medical data [9–11]. In biomedicine, data on the survival times of study participants are often incomplete data. In post-treatment follow-up, some study participants may be found to have died, resulting in accurate survival times being recorded for them. In most studies, however, at least some participants may still be alive when the observation period ends. If this is the case, we cannot obtain the exact survival time for these participants, and we know only that their survival time is longer than the study period. These data can be used to evaluate the treatment effect of drugs by constructing a model of survival function [7, 8].

In clinical trials, a mixture of precise and set-valued medical data may be measured because of the limits of the biotechnologies employed. In frontline chemotherapy for acute lymphoblastic leukemia, minimal residual disease (MRD) is commonly used as a measure of the treatment response after induction, intensification, and consolidation therapy, and it is considered a risk factor with respect to subsequent event-free survival [9–11]. MRD is evaluated by real-time quantitative polymerase chain reaction analysis or flow cytometry or by second-generation sequencing of antigen receptor rearrangements. Different techniques have different sensitivities; for example, flow cytometry can detect 1 leukemic cell in 10000 cells but second-generation sequencing can detect down to 1 cell in 100000 cells. No matter which techniques are used, if the MRD levels are within the range of those techniques, we can obtain precise MRD values; otherwise, i.e., when the MRD is lower than the threshold or undetectable, we can obtain only set-valued data. This is an example of data comprising mixed precise and set-valued measurements.

(iii) Economic systems [12]. In economics, there are many problems relating to mixed data. Tobin [12] studied the problem of consumer demand for household durable goods by collecting data from different families. The collected data were always precise and were recorded as 0 if the family had no consumption. However, no consumption does not mean that there was no demand, and we can know only that the demand of a given family was less than a certain value. Another common problem is determining the actual demand for tickets to an event. If the tickets are not sold out, precise information regarding the demand can be obtained. Otherwise, we know only that the demand was at least equal to the number of tickets sold.

1.2 Related and existing work

To analyze saturated precise or set-valued data, there are three types of approaches that can be applied: (i) using only the precise data; (ii) using only the set-valued data; and (iii) combining all the mixed set-valued and precise data.

For the precise data, many classical identification methods have been developed during the past several decades, such as the maximum likelihood (ML) method [13], the least square (LS) method [14], kernel methods [15, 16], errors-in-variables methods [17, 18], and the blind identification method [19]. In cases

where only the data of the precise part are used, these methods may cause some bias or prove complicated, because the data are truncated and no longer follow the initial distribution. In particular, these methods might not guarantee the accuracy of the estimation, as there are much less precise data than set-valued data.

System identification with set-valued measurements was first investigated by Wang et al. [20]. Consequently, several useful set-valued system identification methods have been proposed, including the empirical measure approach [3, 21, 22], the expectation maximum (EM) algorithm [23–25], and the recursive projection algorithm [26]. Some latest ideas and studies on set-valued identification research can be seen in [27–31]. To use these methods, the identification problem should be solved by regarding the data of the precise part as set-valued data, which leads to the loss of useful information. In addition, the performance of the estimation might be inferior to that of estimations using the precise data.

Combining all the precise and set-valued data is a better approach. The common method is to solve the ML estimation based on the overall likelihood function of a mixed distribution that consists of a discrete point and a continuous variable. Some iterative algorithms, such as the gradient descent method, have been adopted, because the ML estimation problem generally has no closed-form solution.

Fair [32] proposed an iterative algorithm that involves more iterations but less computation time; however, the algorithm must choose suitable initial values. Heckman [33] proposed a two-stage estimation algorithm that first computed the initial parameter estimation with the set-valued data and then combined the precise data to obtain the final parameter estimation by the LS method. However, this method is less effective than the ML estimation, despite its simple calculation process. Although these methods provide feasible solutions, they focus mainly on the convergence of the algorithms, and the identification effects, such as the asymptotic efficiency, might not be guaranteed. Thus, there is an urgent need to establish better identification methods.

1.3 Ideas and contributions

This paper aims to develop a feasible and unified method based on the mixed precise and set-valued data. First, suitable identification methods for each type of data are selected. Then, an estimation fusion algorithm is constructed. The benefit of this method is that it deals with each type of data separately and then combines the two estimations in a manner, such as linear combination, that overcomes the difficulty in solving the identification problem based on the mixed distributions. In other words, different types of data should be classified for estimation, because the information they carry is different.

Finally, and most importantly, how should suitable weights be chosen for two different identification results? In general, the Cramér-Rao (CR) lower bound expresses a lower bound of the variance of the parameter estimation. An intuitive but not gratuitous view is that we should give greater weight to the estimator with the smaller variance as it can provide more accurate information. Furthermore, the CR lower bound can be divided into two independent parts that are exactly associated with the variances of the two estimators for the precise data and the set-valued data, respectively. This encourages us to construct the proposed asymptotically efficient algorithm.

To establish the above concept, we consider a linear system with the mixed measurements under periodic inputs. ML estimators based on the whole system, the precise subsystem, and the set-valued subsystem are obtained. Then, we mainly analyze the convergence properties of the ML estimator and the CR lower bound-based algorithm (CRBA) in terms of consistency and asymptotic efficiency. Following a similar concept, we propose an iterative algorithm to solve the identification problem under general inputs.

The contributions of this paper can be summarized as follows.

- A new system identification problem is considered under saturated precise or set-valued data, which can be formulated as an optimal data-integration problem under combined continuous density and discrete cumulative distribution functions. The problem is new and its treatment is beyond the traditional identification framework that deals with only probability density functions or set-valued one that involves probability cumulative functions.

- This paper fills the gap between the mixed data and existing methodologies, and as far as we know, is the first unified framework in dealing with such problems rigorously and comprehensively. New algorithms are developed and their properties are established.

- The CR lower bound based on the mixed data is established. Its structure consists of two parts with respect to the precise data and set-valued data, respectively, which enlightens us to construct a CRBA.

Under periodic inputs, the CRBA is proved to be strongly consistent and asymptotically achieve the CR lower bound, hence having efficient property. Under general inputs, the CRBA is constructed to combine the EM algorithm for the set-valued data and the gradient descent algorithm for the precise data.

The rest of the paper is organized as follows. Section 2 introduces the system with mixed precise and set-valued measurements. Section 3 constructs different estimators of the system parameters under periodic inputs. Section 4 derives the convergence properties of the CRBA and the JMLE based on the whole system in Section 3 in terms of consistency and asymptotic efficiency. Section 5 proposes an iterative algorithm based on the CRBA under general inputs. Section 6 discusses the difference between the CRBA and the ML estimator. Section 7 demonstrates some results obtained through extensive numerical simulations. Section 8 summarizes the paper and discusses related future work.

2 Problem formulation

In this section, the general form of a system with mixed precise and set-valued measurements is given first. A linear system with mixed precise and set-valued measurements is then introduced, and the corresponding system identification problem is formulated.

A general form of a system with precise and set-valued measurements can be described as follows:

$$\begin{cases} y_k = f(\phi_k, \theta) + e_k, \\ s_k = \begin{cases} y_k, & \text{if } y_k \in A, \\ Q(y_k), & \text{otherwise}, \end{cases} \end{cases} \tag{1}$$

where $k = 1, 2, \ldots, N$, ϕ_k is the system input, y_k is the unobserved system output, s_k is the measurement of y_k, and function Q is a data quantizer or sensor. s_k denotes precise data if y_k belongs to set A and otherwise denotes set-valued data. e_k is the system noise with zero mean and standard derivation σ, θ is the unknown constant parameter vector, and function $f(\cdot)$ is a deterministic function of ϕ_k and θ, which represents the model structure.

Similar to system (1), a linear system with precise and set-valued measurements is described as follows:

$$\begin{cases} y_k = \phi_k^{\mathrm{T}}\theta + e_k, \\ s_k = \begin{cases} y_k, & \text{if } y_k > c, \\ c, & \text{otherwise}, \end{cases} \end{cases} \quad k = 1, 2, \ldots, N, \tag{2}$$

where $\theta \in \mathbb{R}^n$ is the system parameter vector, $\phi_k \in \mathbb{R}^n$ is the system input, $y_k \in \mathbb{R}$ is the unobserved system output, $c \in \mathbb{R}$ is a given set-valued sensor threshold, s_k is the measurement of y_k, which is precise data if $y_k > c$, and set-valued data otherwise.

The system identification problem is to estimate the parameter vector θ based on $\{\phi_k, s_k, 1 \leqslant k \leqslant N\}$, namely,

$$\{\phi_k, s_k, 1 \leqslant k \leqslant N\} \longrightarrow \{\theta\}.$$

In consideration of the complexity of system (2), the following assumptions are made.

Assumption 1. The system noise $\epsilon_N = \{e_1, e_2, \ldots, e_N\}$ is an independent and identically distributed (i.i.d) stochastic sequence that follows a standard normal distribution, namely, $\sigma = 1$.

Remark 1. For simplicity, we assume that $\sigma = 1$ and is known as above. In fact, if the value of σ is unknown, we can estimate it from the set-valued data [34].

Assumption 2. The system inputs follow the condition that

$$\limsup_N \frac{1}{N} \sum_{k=1}^{N} \phi_k \phi_k^{\mathrm{T}} > \delta I, \tag{3}$$

where $\delta > 0$ and I is the identity matrix.

Remark 2. Assumption 2 is the mathematical description of persistent excitation condition, which is a typical assumption in the research of system identification [2].

In the following of this paper, we will propose an identification algorithm under mixed precise and set-valued data. The CR lower bound based on the mixed precise and set-valued data will be calculated

and a CRBA will be proposed. Under periodic inputs, the algorithm is proved to be strongly consistent and asymptotically efficient. Under general inputs, an iterative algorithm based on the CRBA will be constructed.

Since the CR lower bounds are independent of algorithms and it has always been an important index to evaluate problem complexity, achieving the CR lower bounds is an ultimate pursuit of identification accuracy in all identification problems. Since this data-fusion problem is new, our results are the first in this new field, optimal in its mean-square convergence rates, and fundamental.

3 System identification under periodic inputs

In this section, several identification methods for system (2) based on the novel ML estimation are discussed under periodic inputs.

Assumption 3. The system inputs $\Phi_N = \{\phi_1, \phi_2, \ldots, \phi_N\}$ have an n-periodic form, that is, $\forall k \geqslant 1, \phi_k = \phi_{k+n}$. The data size N is divisible by n, which satisfies $N/n = \bar{N}$.

For convenience, the measurements $\Omega_N = \{s_1, \ldots, s_N\}$ are subdivided into n separate groups with respect to periodic inputs, namely, $\{s_{11}, \ldots, s_{1\bar{N}}\}, \ldots, \{s_{n1}, \ldots, s_{n\bar{N}}\}$, where s_{ij} denotes the measurement $s_{i+(j-1)n}$ under the system input $\phi_{i+(j-1)n}$ $(= \phi_i)$, $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant \bar{N}$.

Assumption 4. For any group of measurements $\{s_{i1}, \ldots, s_{i\bar{N}}\}$, $1 \leqslant i \leqslant n$, there are both precise and set-valued data.

Remark 3. Assumption 4 is necessary; otherwise, two situations will occur. In the first situation, where there is only pure accurate data, the algorithm for accurate data can be used. However, in the second situation, where there is only saturation data, the algorithm for set-valued data will not work, because the information from the set-valued output is insufficiently rich.

Here, we define an input matrix $\Psi_N = (\phi_1, \phi_2, \ldots, \phi_n)^{\mathrm{T}}$ and a vector $\psi_N = (\psi_1, \ldots, \psi_n)^{\mathrm{T}} = \Psi_N \theta$. Under Assumption 2, Ψ_N is guaranteed to be invertible. If we can obtain the estimation $\hat{\psi}_N$ of the parameter ψ, then the estimation $\hat{\theta}_N$ of the parameter θ will be

$$\hat{\theta}_N = \Psi_N^{-1} \hat{\psi}_N. \tag{4}$$

In this sense, the estimation problem of the parameter θ is turned into the estimation problem for the parameter ψ.

Remark 4. In Assumption 3, the number of periods is equal to the dimension n of parameter θ, but this is not necessary. In cases where the number of periods is larger than n, we can construct the algorithm with the invertible matrix $\Psi_N^{\mathrm{T}} \Psi_N$.

In this paper, most symbols are related to the data size N, such as the log-likelihood function $l_N(\theta)$, input matrix Ψ_N, and the related ML estimations $\hat{\theta}_N$ and $\hat{\psi}_N$. For simplicity, let $l(\theta), \Psi, \hat{\theta}$, and $\hat{\psi}$ denote the corresponding symbols, respectively.

3.1 Joint identification

For system (2), the probability density distribution of s_k can be written as

$$f_s(x) = [F(c - \phi_k^{\mathrm{T}} \theta)]^{\mathbb{I}_{\{x=c\}}} [f(x - \phi_k^{\mathrm{T}} \theta)]^{\mathbb{I}_{\{x \neq c\}}}, \tag{5}$$

where $f(\cdot)$ and $F(\cdot)$ are the probability density function and cumulative distribution function, respectively, of the standard normal random variable; c is the threshold and x is the value of precise system output.

Remark 5. Eq. (5) shows a mixed distribution that involves mixed continuous and discrete random variables. This mixture of different types of random variables introduces highly non-trivial technical challenges such as how to group and mix the data, and how to develop optimal weights on data to reduce identification errors.

Proposition 1. For system (2), under Assumptions 1–4, the ML estimator for the parameter θ, based on the joint likelihood method is

$$\hat{\theta}^{\mathrm{J}} = \Psi^{-1} \hat{\psi}^{\mathrm{J}}, \tag{6}$$

where $\hat{\psi}^{\mathrm{J}} = [\hat{\psi}_1^{\mathrm{J}}, \ldots, \hat{\psi}_n^{\mathrm{J}}]^{\mathrm{T}}$, $\hat{\psi}_i^{\mathrm{J}} = G^{-1}(\frac{1}{\bar{N}} \sum_{j=1}^{\bar{N}} s_{ij})$, $1 \leqslant i \leqslant n$, and $G(x) = (x - c)F(x - c) + f(x - c) + c$, which is a monotonous continuous function.

Proof. The joint log-likelihood function that represents the overall probability of the measurements Ω_N given the inputs Φ_N and the parameter θ is

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{\bar{N}} \log \mathbb{P}(s_{ij}|\phi_i,\theta) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\bar{N}} \Big[\log F(c-\psi_i)\cdot \mathbb{I}_{\{s_{ij}=c\}} + \log[f(s_{ij}-\psi_i)]\mathbb{I}_{\{s_{ij}\neq c\}}\Big]. \end{aligned} \tag{7}$$

The ML estimation denoted by $\hat{\theta}^{\mathrm{J}}$ is the parameter that maximizes the log-likelihood function $l(\theta)$,

$$\hat{\theta}^{\mathrm{J}} = \arg\max_{\theta} l(\theta). \tag{8}$$

We first compute the gradient vector of $l(\theta)$,

$$\nabla l(\theta) = \sum_{j=1}^{\bar{N}} [\phi_1,\ldots,\phi_n]J_j = \Psi^{\mathrm{T}} J, \tag{9}$$

where $J = J_1 + \cdots + J_{\bar{N}}$ and $J_j, 1 \leqslant j \leqslant \bar{N}$, is an n-dimensional vector defined as

$$J_j = \begin{bmatrix} \frac{-f(c-\psi_1)}{F(c-\psi_1)}\mathbb{I}_{\{s_{1j}=c\}} + (s_{1j}-\psi_1)\mathbb{I}_{\{s_{1j}\neq c\}} \\ \vdots \\ \frac{-f(c-\psi_n)}{F(c-\psi_n)}\mathbb{I}_{\{s_{nj}=c\}} + (s_{nj}-\psi_n)\mathbb{I}_{\{s_{nj}\neq c\}} \end{bmatrix}. \tag{10}$$

Because Ψ^{T} is an invertible matrix, it is easy to obtain from $\nabla l(\theta) = 0$ that $J = 0$. That is, each component of J equals 0. For $1 \leqslant i \leqslant n$, it follows that

$$\sum_{j=1}^{\bar{N}} \frac{-f(c-\psi_i)}{F(c-\psi_i)}\mathbb{I}_{\{s_{ij}=0\}} + (s_{ij}-\psi_i)\mathbb{I}_{\{s_{ij}\neq 0\}} = 0,$$

which indicates that it can be regarded as a one-dimensional system with a single parameter ψ_i and constant input 1. Given parameter ψ_i, the measurements $\{s_{i1},\ldots,s_{i\bar{N}}\}$ are samples with the probability density

$$f_s(x) = [F(c-\psi_i)]^{\mathbb{I}_{\{x=0\}}}[f(x-\psi_i)]^{\mathbb{I}_{\{x\neq 0\}}} = [F(c-\psi_i)]^{\mathbb{I}_{\{x=0\}}}\left[\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\psi_i)^2}{2}}\right]^{\mathbb{I}_{\{x\neq 0\}}}.$$

Suppose that the number of precise data $\sum_{j=1}^{\bar{N}}\mathbb{I}_{\{s_{ij}\neq 0\}} = \bar{N}_{i1}$, $1 \leqslant i \leqslant n$; thus

$$\begin{aligned} &-(\bar{N}-\bar{N}_{i1})\frac{f(c-\psi_i)}{F(c-\psi_i)} + \sum_{j=1}^{\bar{N}_{i1}} s_{ij} - \bar{N}_{i1}\psi_i = 0, \\ &\psi_i + \frac{\bar{N}-\bar{N}_{i1}}{\bar{N}_{i1}}\frac{f(c-\psi_i)}{F(c-\psi_i)} = \frac{1}{\bar{N}_{i1}}\sum_{j=1}^{\bar{N}_{i1}} s_{ij}. \end{aligned} \tag{11}$$

Then, the ML estimation of parameter ψ_i is the solution of (11). Let us now look at the moment estimation of $\{\psi_i, 1 \leqslant i \leqslant n\}$. The expectation of $\{s_{ij}, 1 \leqslant j \leqslant \bar{N}\}$ is

$$\begin{aligned} \mathbb{E}s_{ij} &= \mathbb{E}[y_{i+(j-1)\bar{N}}|y_{i+(j-1)\bar{N}} > c]\mathbb{P}(y_{i+(j-1)\bar{N}} > c) + c\cdot\mathbb{P}(y_{i+(j-1)\bar{N}} \leqslant c) \\ &= \mathbb{E}[y_{i+(j-1)\bar{N}}|y_{i+(j-1)\bar{N}} > c]F(\psi_i - c) + cF(c-\psi_i) \\ &= (\psi_i - c)F(\psi_i - c) + f(\psi_i - c) + c, \end{aligned} \tag{12}$$

where $\mathbb{E}$ and $\mathbb{P}$ are the expectation and probability, respectively. Because $G'(x) = F(\theta - c) > 0$ holds for $x \in \mathbb{R}$, the first-moment estimation of ψ_i is defined as

$$\hat{\psi}_i^{\mathrm{J}} = G^{-1}\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}} s_{ij}\right). \tag{13}$$

Then $P(s_{ij} = 0) = F(c - \psi_i), \forall 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant \bar{N}$, and it can be inferred that

$$\bar{N}_{i1} \to \bar{N}(1 - F(c - \psi_i)), \quad \bar{N} \to \infty, \tag{14}$$

which is put into (11) to obtain (13). This means that the ML estimation and first-moment estimation of $\{\psi_i, 1 \leqslant i \leqslant n\}$ are equivalent in the sense that $\bar{N} \to \infty$. Thus, we obtain the estimator $\hat{\psi}^{\mathrm{J}}$ for the parameter vector ψ and then the estimator $\hat{\theta}^{\mathrm{J}}$ in (6).

Remark 6. For convenience, we will use "JMLE" to represent the identification method based on the joint likelihood function.

3.2 Subsystem identification

Without loss of generality, for a group of measurements $\{s_{i1}, \ldots, s_{i\bar{N}}\}$, $1 \leqslant i \leqslant n$, suppose that the corresponding former $\bar{N}_{i1}$ measurements are precise data. Then, the measurements $\{s_{i1}, \ldots, s_{i\bar{N}}\}$ can be divided into two parts consisting of precise data $\{s_{ij}^{(1)}, 1 \leqslant j \leqslant \bar{N}_{i1}\}$ and set-valued data $\{s_{ij}^{(2)}, 1 \leqslant j \leqslant \bar{N}\} = \{1, \ldots, 1, 0, \ldots, 0\}$, respectively, and can be regarded as the measurements of two subsystems of system (2). Next, we will determine the estimators of the parameter θ in terms of the precise subsystem and the set-valued subsystem.

3.2.1 *Precise subsystem identification*

The precise subsystem collects only precise measurements. For any $i \leqslant n, j \leqslant \bar{N}_{i1}$, the probability density function of the measurement $s_{ij}^{(1)}$ is as follows:

$$f_{s_{ij}^{(1)}}(x|\phi_i, \theta) = \frac{f(x - \phi_i^{\mathrm{T}}\theta)}{1 - F(c - \phi_i^{\mathrm{T}}\theta)}. \tag{15}$$

Similar to the proof of Proposition 1, we have Proposition 2.

Proposition 2. For system (2), under Assumptions 1–4, the ML estimator of the parameter θ, based on the precise measurements, is

$$\hat{\theta}^{\mathrm{P}} = \Psi^{-1}\hat{\psi}^{\mathrm{P}}, \tag{16}$$

where $\hat{\psi}^{\mathrm{P}} = [\hat{\psi}_1^{\mathrm{P}}, \ldots, \hat{\psi}_n^{\mathrm{P}}]^{\mathrm{T}}$, $\hat{\psi}_i^{\mathrm{P}} = H^{-1}(\frac{1}{N_{i1}}\sum_{j=1}^{N_{i1}} s_{ij}^{(1)})$, $1 \leqslant i \leqslant n$, and $H(x) = x + \frac{f(x-c)}{F(x-c)}$, which is a monotonous continuous function.

3.2.2 *Set-valued subsystem identification*

For system identification with periodic inputs and set-valued measurements, the empirical measure approach proposed in [3] is actually equivalent to the ML estimator or moment estimator.

Proposition 3. For system (2), under Assumptions 1–4, the ML estimator of the parameter θ, based on the set-valued measurements, is

$$\hat{\theta}^{\mathrm{B}} = \Psi^{-1}\hat{\psi}^{\mathrm{B}}, \tag{17}$$

where $\hat{\psi}^{\mathrm{B}} = [\hat{\psi}_1^{\mathrm{B}}, \ldots, \hat{\psi}_n^{\mathrm{B}}]^{\mathrm{T}}$, $\hat{\psi}_i^{\mathrm{B}} = F^{-1}(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}} s_{ij}^{(2)}) + c = F^{-1}(\frac{N_{i1}}{\bar{N}}) + c$, $1 \leqslant i \leqslant n$.

Proof. The proof can be derived according to the method of Wang et al. [3].

4 Convergence property

To evaluate the different identification results described above, we will take the strong consistency and the asymptotic efficiency as the evaluation metrics for the convergence property of each estimator.

4.1 Strong consistency

Theorem 1. For system (2) under periodic inputs, the JMLE estimator $\hat{\theta}^{\mathrm{J}}$ in (6) is asymptotically unbiased and strongly consistent in the sense that

$$\begin{aligned}\hat{\theta}^{\mathrm{J}} &\to \theta \quad \text{w.p.1}, \quad N \to \infty,\\ \hat{\theta}^{\mathrm{P}} &\to \theta \quad \text{w.p.1}, \quad N \to \infty,\\ \hat{\theta}^{\mathrm{B}} &\to \theta \quad \text{w.p.1}, \quad N \to \infty.\end{aligned}$$

Proof. In the previous analysis, each component $\hat{\psi}_i^{\mathrm{J}}, 1 \leqslant i \leqslant n$, of $\hat{\psi}^{\mathrm{J}}$ can be regarded as a function with respect to the first-order sample moment of $\{s_{ij}, 1 \leqslant j \leqslant \bar{N}\}$. The same applies to $\hat{\psi}^{\mathrm{B}}$ and $\hat{\psi}^{\mathrm{P}}$. As a result of the continuity of the functions $G^{-1}(\cdot), F^{-1}(\cdot)$, and $H^{-1}(\cdot)$, $\hat{\psi}_i^{\mathrm{J}}, \hat{\psi}_i^{\mathrm{B}}$, and $\hat{\psi}_i^{\mathrm{P}}, 1 \leqslant i \leqslant n$ are all strongly consistent estimators of ψ_i according to the strong law of large numbers [35]. Thus, $\hat{\psi}^{\mathrm{J}}, \hat{\psi}^{\mathrm{B}}$, and $\hat{\psi}^{\mathrm{P}}$ are all strongly consistent estimators of ψ. Furthermore, because linear transformations do not affect the convergence of the estimators, $\hat{\theta}^{\mathrm{J}}, \hat{\theta}^{\mathrm{B}}$, and $\hat{\theta}^{\mathrm{P}}$ are all strongly consistent.

4.2 Asymptotic efficiency

Theorem 2. For system (2) under periodic inputs, the CR lower bound of the parameter estimation is

$$\Sigma_{\mathrm{CR}} = \frac{1}{\bar{N}} \Psi^{-1} \Sigma_{\psi} \Psi^{-\mathrm{T}}, \tag{18}$$

where Σ_{ψ} is a diagonal matrix and its ith $(1 \leqslant i \leqslant n)$ diagonal element is

$$\begin{aligned}\Sigma_{\psi}[i,i] &= \left[\frac{f^2(\psi_i - c)}{1 - F(\psi_i - c)} + F(\psi_i - c) - f(\psi_i - c)(\psi_i - c)\right]^{-1}\\ &= \left[\frac{f^2(\psi_i - c)}{(1 - F(\psi_i - c))F(\psi_i - c)} + \frac{f^2(\psi_i - c)}{F(\psi_i - c)} + F(\psi_i - c) - f(\psi_i - c)(\psi_i - c)\right]^{-1}\\ &= \left[\frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{(1 - F(\phi_i^{\mathrm{T}}\theta - c))F(\phi_i^{\mathrm{T}}\theta - c)} + \frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{F(\phi_i^{\mathrm{T}}\theta - c)} + F(\phi_i^{\mathrm{T}}\theta - c) - f(\phi_i^{\mathrm{T}}\theta - c)(\phi_i^{\mathrm{T}}\theta - c)\right]^{-1}.\end{aligned}$$

Proof. Notice that the CR low bound is based on the likelihood function (7). Then its second-order derivative is

$$\frac{\partial l^2(\theta)}{\partial \theta^2} = \sum_{i=1}^{n}\sum_{j=1}^{\bar{N}}\left[\frac{f(c - \phi_i^{\mathrm{T}}\theta)(c - \phi_i^{\mathrm{T}}\theta)F(c - \phi_i^{\mathrm{T}}\theta) + f^2(c - \phi_i^{\mathrm{T}}\theta)}{-F^2(c - \phi_i^{\mathrm{T}}\theta)}\mathbb{I}_{\{s_{ij}=c\}}\phi_i\phi_i^{\mathrm{T}} - \mathbb{I}_{\{s_{ij}\neq c\}}\phi_i\phi_i^{\mathrm{T}}\right].$$

Then, because for $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant \bar{N}$,

$$\mathbb{E}\mathbb{I}_{\{s_{ij}=c\}} = F(c - \phi_i^{\mathrm{T}}\theta) = 1 - F(\phi_i^{\mathrm{T}}\theta - c),$$

the CR lower bound is

$$\begin{aligned}\Sigma_{\mathrm{CR}} &= -\left(\mathbb{E}\frac{\partial l^2(\theta)}{\partial \theta^2}\right)^{-1}\\ &= \frac{1}{\bar{N}}\left[\sum_{i=1}^{n}\left[\frac{f(c - \phi_i^{\mathrm{T}}\theta)(c - \phi_i^{\mathrm{T}}\theta)F(c - \phi_i^{\mathrm{T}}\theta) + f^2(c - \phi_i^{\mathrm{T}}\theta)}{F(c - \phi_i^{\mathrm{T}}\theta)}\phi_i\phi_i^{\mathrm{T}} + F(\phi_i^{\mathrm{T}}\theta - c)\phi_i\phi_i^{\mathrm{T}}\right]\right]^{-1}\\ &= \frac{1}{\bar{N}}\left[\sum_{i=1}^{n}\left[-f(\phi_i^{\mathrm{T}}\theta - c)(\phi_i^{\mathrm{T}}\theta - c) + \frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{1 - F(\phi_i^{\mathrm{T}}\theta - c)} + F(\phi_i^{\mathrm{T}}\theta - c)\right]\phi_i\phi_i^{\mathrm{T}}\right]^{-1}\\ &= \frac{1}{\bar{N}}\left[[\phi_1, \ldots, \phi_n]\,\Sigma_{\psi}^{-1}[\phi_1, \ldots, \phi_n]^{\mathrm{T}}\right]^{-1}\\ &= \frac{1}{\bar{N}}\Psi^{-1}\Sigma_{\psi}\Psi^{-\mathrm{T}}.\end{aligned}$$

Remark 7. In fact, it can be seen that the CR lower bound consists of two parts that are associated with the respective covariances of two estimators based on the subsystem identification results. By (18), we define

$$\Sigma_{\mathrm{B}} = \frac{1}{\bar{N}}\Psi^{-1}\Sigma_{\mathrm{B}}^{\psi}\Psi^{-\mathrm{T}}, \tag{19}$$

$$\Sigma_{\mathrm{P}} = \frac{1}{\bar{N}}\Psi^{-1}\Sigma_{\mathrm{P}}^{\psi}\Psi^{-\mathrm{T}}, \tag{20}$$

where $\Sigma_{\mathrm{B}}^{\psi}$ and $\Sigma_{\mathrm{P}}^{\psi}$ are diagonal and the corresponding ith $(1 \leqslant i \leqslant n)$ diagonal elements are

$$\Sigma_{\mathrm{B}}^{\psi}[i,i] = \left[\frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{(1 - F(\phi_i^{\mathrm{T}}\theta - c))F(\phi_i^{\mathrm{T}}\theta - c)}\right]^{-1}, \tag{21}$$

$$\Sigma_{\mathrm{P}}^{\psi}[i,i] = \left[\frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{F(\phi_i^{\mathrm{T}}\theta - c)} + F(\phi_i^{\mathrm{T}}\theta - c) - f(\phi_i^{\mathrm{T}}\theta - c)(\phi_i^{\mathrm{T}}\theta - c)\right]^{-1}. \tag{22}$$

Thus, it follows that

$$\Sigma_{\mathrm{CR}}^{-1} = \Sigma_{\mathrm{B}}^{-1} + \Sigma_{\mathrm{P}}^{-1}. \tag{23}$$

The special structure of the CR lower bound is exactly the reason that enables the optimal algorithm to be found.

Theorem 3. For system (2) under periodic inputs, the JMLE estimator $\hat{\theta}^{\mathrm{J}}$ is not asymptotically efficient; that is, the CR lower bound Σ_{CR} cannot be achieved:

$$N[\mathbb{E}(\hat{\theta}^{\mathrm{J}} - \theta)(\hat{\theta}^{\mathrm{J}} - \theta)^{\mathrm{T}} - \Sigma_{\mathrm{CR}}] \nrightarrow 0, \quad N \to \infty. \tag{24}$$

Proof. First, we compute the asymptotic distribution of the estimator $\hat{\psi}^{\mathrm{J}}$. For $1 \leqslant i \leqslant n$, measurements $\{s_{ij}, 1 \leqslant j \leqslant \bar{N}\}$ can be considered as i.i.d. samples from a one-dimensional system with a single parameter ψ_i and constant input 1. In terms of (12), the expectation of s_{ij} is

$$\mathbb{E}s_{ij} = G(\psi_i) = (\psi_i - c)F(\psi_i - c) + f(\psi_i - c) + c.$$

Then, the variance of s_{ij} is

$$\begin{aligned} V_{s_i} &= \mathbb{E}[s_{ij}]^2 - [\mathbb{E}s_{ij}]^2 \\ &= F(\psi_i - c)[1 - F(\psi_i - c)]\left[\psi_i - c + \frac{f(\psi_i - c)}{F(\psi_i - c)}\right]^2 + F(\psi_i - c) - \frac{f^2(\psi_i - c)}{F(\psi_i - c)} - (\psi_i - c)f(\psi_i - c). \end{aligned}$$

According to the central-limit theorem, we know that

$$\sqrt{\bar{N}}\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}} s_{ij} - \mathbb{E}s_{ij}\right) \to N(0, V_{s_i}), \quad \bar{N} \to \infty.$$

From Subsection 3.1, the existence and continuity of the inverse function $G^{-1}(\cdot)$ are guaranteed; thus,

$$\begin{aligned} \sqrt{\bar{N}}(\hat{\psi}_i^{\mathrm{J}} - \psi_i) &= \sqrt{\bar{N}}\left(G^{-1}\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}} s_{ij}\right) - \psi_i\right) \\ &\to N\left(0, \frac{V_{s_i}}{F^2(\psi_i - c)}\right), \quad \bar{N} \to \infty, \end{aligned}$$

where the derivative of $G^{-1}(\cdot)$ is $1/F(x - c)$. Therefore, the asymptotic distribution of the estimator $\hat{\psi}^{\mathrm{J}}$ is

$$\sqrt{\bar{N}}(\hat{\psi}^{\mathrm{J}} - \psi) \to N(0, \Sigma_J^{\psi}), \quad \bar{N} \to \infty,$$

where $\Sigma_{\mathrm{J}}^{\psi}$ is a diagonal matrix and its ith $(1 \leqslant i \leqslant n)$ diagonal element is $\Sigma_{\mathrm{J}}^{\psi}[i,i] = \frac{V_{s_i}}{F^2(\psi_i - c)}$. From (4), the asymptotic distribution of the estimator $\hat{\theta}^{\mathrm{J}}$ is

$$\sqrt{\bar{N}}(\hat{\theta}^{\mathrm{J}} - \theta) \to N(0, \Psi^{-1}\Sigma_{\mathrm{J}}^{\psi}\Psi^{-\mathrm{T}}), \quad \bar{N} \to \infty.$$

Thus, we have

$$\mathbb{E}(\hat{\theta}^{\mathrm{J}} - \theta)(\hat{\theta}^{\mathrm{J}} - \theta)^{\mathrm{T}} \to \frac{1}{\bar{N}}\Psi^{-1}\Sigma_{\mathrm{J}}^{\psi}\Psi^{-\mathrm{T}}. \tag{25}$$

Comparing this with (18), it can be concluded that $\hat{\theta}^{\mathrm{J}}$ cannot asymptotically achieve the CR lower bound Σ_{CR}, which implies (24).

4.3 CRBA

Theorem 4. For system (2) under periodic inputs, we define A and B as follows:

$$A = \frac{1}{\bar{N}}\Sigma_{\mathrm{B}}^{-1}, \qquad B = \frac{1}{\bar{N}}\Sigma_{\mathrm{P}}^{-1},$$

where Σ_{B} and Σ_{P} are defined in (19)–(22). Then, the estimator is asymptotically efficient if

$$\hat{\theta}^{\mathrm{C}} = \Pi_{\mathrm{B}}\hat{\theta}^{\mathrm{B}} + \Pi_{\mathrm{P}}\hat{\theta}^{\mathrm{P}}, \tag{26}$$

where

$$\Pi_{\mathrm{B}} = (A+B)^{-1}A, \quad \Pi_{\mathrm{P}} = (A+B)^{-1}B. \tag{27}$$

That is, it can asymptotically achieve the CR lower bound Σ_{CR}:

$$N[\mathbb{E}(\hat{\theta}^{\mathrm{C}} - \theta)(\hat{\theta}^{\mathrm{C}} - \theta)^{\mathrm{T}} - \Sigma_{\mathrm{CR}}] \to 0, \quad N \to \infty. \tag{28}$$

Proof. First, we will prove that the covariances of the estimators $\hat{\theta}^{\mathrm{B}}$ and $\hat{\theta}^{\mathrm{P}}$ can asymptotically achieve the two parts Σ_{B} and Σ_{P}, respectively, of the CR lower bound by the same method as that in Theorem 3.

Step 1. **The covariance of $\hat{\theta}^{\mathrm{P}}$ $\to \boldsymbol{\Sigma}_{\mathrm{P}}$.** For any $1 \leqslant i \leqslant n$, a group of measurements $\{s_{ij}^{(1)}, 1 \leqslant j \leqslant \bar{N}_{i1}\}$ can be considered as i.i.d. samples from a one-dimensional precise subsystem with a single parameter ψ_i and constant input 1. The second-order origin moment of $s_{ij}^{(1)}$ is

$$\mathbb{E}[s_{ij}^{(1)}]^2 = \mathbb{E}(y^2_{i+(j-1)\bar{N}_{i1}} | y_{i+(j-1)\bar{N}_{i1}} > c) = \frac{f(\psi_i - c)(\psi_i + c) + F(\psi_i - c)(1 + \psi_i^2)}{F(\psi_i - c)}.$$

Then, the variance of $s_{ij}^{(1)}$ is

$$V_{s_i^{(1)}} = \mathbb{E}[s_{ij}^{(1)}]^2 - [\mathbb{E}s_{ij}^{(1)}]^2 = \left[1 - \frac{f(\psi_i - c)(\psi_i - c)F(\psi_i - c) + f^2(\psi_i - c)}{F^2(\psi_i - c)}\right].$$

According to the central-limit theorem, we know that

$$\sqrt{\bar{N}_{i1}}\left(\frac{1}{\bar{N}_{i1}}\sum_{j=1}^{\bar{N}_{i1}} s_{ij}^{(1)} - \mathbb{E}s_{ij}^{(1)}\right) \to N(0, V_{s_i^{(1)}}), \quad \bar{N}_{i1} \to \infty.$$

Thus, we have

$$\sqrt{\bar{N}_{i1}}(\hat{\psi}_i^{\mathrm{P}} - \psi_i) = \sqrt{\bar{N}_{i1}}\left(H^{-1}\left(\frac{1}{\bar{N}_{i1}}\sum_{j=1}^{\bar{N}_{i1}} s_{ij}^{(1)}\right) - \psi_i\right) \to N\left(0, V_{s_i^{(1)}}^{-1}\right), \quad \bar{N}_{i1} \to \infty,$$

where the derivative of $H^{-1}(\cdot)$ is

$$\left[1 - \frac{f(x-c)(x-c)F(x-c) + f^2(x-c)}{F^2(x-c)}\right]^{-2}.$$

Therefore, the asymptotic distribution of the estimator $\hat{\psi}^{\mathrm{P}}$ is

$$\hat{\psi}^{\mathrm{P}} - \psi \to N(0, \Sigma_{P_0}^{\psi}), \quad \bar{N}_{i1} \to \infty,$$

where $\Sigma_{\mathrm{P}}^{\psi}$ is a diagonal matrix and its ith $(1 \leqslant i \leqslant n)$ diagonal element is $\Sigma_{P_0}^{\psi}[i,i] = \frac{1}{\bar{N}_{i1} V_{s_i^{(2)}}}$.

As $\bar{N} \to \infty$, $\bar{N}_{i1} \to \bar{N} F(\psi_i - c), 1 \leqslant i \leqslant n$, $\Sigma_{P_0}^{\psi}$ can be rewritten as $\frac{1}{\bar{N}} \Sigma_{\mathrm{P}}^{\psi}$ in this sense,

$$\Sigma_{P_0}^{\psi} \to \frac{1}{\bar{N}} \begin{bmatrix} \frac{1}{F(\psi_1 - c) V_{s_1^{(1)}}} & & \\ & \ddots & \\ & & \frac{1}{F(\psi_n - c) V_{s_n^{(1)}}} \end{bmatrix} = \frac{1}{\bar{N}} \Sigma_{\mathrm{P}}^{\psi},$$

where $\Sigma_{\mathrm{P}}^{\psi}$ is defined in (22). From (16), the asymptotic distribution of the estimator $\hat{\theta}^{\mathrm{P}}$ is

$$\hat{\theta}^{\mathrm{P}} - \theta \to N\left(0, \frac{1}{\bar{N}} \Psi^{-1} \Sigma_{\mathrm{P}}^{\psi} \Psi^{-\mathrm{T}}\right), \quad \bar{N} \to \infty.$$

Then, we have

$$\mathbb{E}(\hat{\theta}^{\mathrm{P}} - \theta)(\hat{\theta}^{\mathrm{P}} - \theta)^{\mathrm{T}} \to \Sigma_{\mathrm{P}}, \quad \bar{N} \to \infty. \tag{29}$$

Step 2. **The covariance of $\hat{\theta}^{\mathrm{B}}$ $\to \boldsymbol{\Sigma}_{\mathrm{B}}$.** The asymptotic distribution of $\hat{\psi}^{\mathrm{B}}$ is

$$\sqrt{\bar{N}}(\hat{\psi}^{\mathrm{B}} - \psi) \to N(0, \Sigma_{\mathrm{B}}^{\psi}), \quad \bar{N} \to \infty,$$

where $\Sigma_{\mathrm{B}}^{\psi}$ is defined in (21). From (17), the asymptotic distribution of the estimator $\hat{\theta}^{\mathrm{B}}$ is

$$\sqrt{\bar{N}}(\hat{\theta}^{\mathrm{B}} - \theta) \to N(0, \Psi^{-1} \Sigma_{\mathrm{B}}^{\psi} \Psi^{-\mathrm{T}}), \quad \bar{N} \to \infty.$$

Then, we have

$$\mathbb{E}(\hat{\theta}^{\mathrm{B}} - \theta)(\hat{\theta}^{\mathrm{B}} - \theta)^{\mathrm{T}} \to \Sigma_{\mathrm{B}}, \quad \bar{N} \to \infty. \tag{30}$$

Step 3. **Construct the optimal algorithm**. In terms of Steps 1 and 2, we have

$$A = \frac{1}{\bar{N}} \Sigma_{\mathrm{B}}^{-1} = \Psi^{\mathrm{T}} [\Sigma_{\mathrm{B}}^{\psi}]^{-1} \Psi,$$
$$B = \frac{1}{\bar{N}} \Sigma_{\mathrm{P}}^{-1} = \Psi^{\mathrm{T}} [\Sigma_{\mathrm{P}}^{\psi}]^{-1} \Psi.$$

Because the CR lower bound is the minimum variance of all (asymptotically) unbiased estimators, the optimality is guaranteed if we can prove that $\hat{\theta}^{\mathrm{CR}}$ can asymptotically achieve the Σ_{CR}. Thus, we need to compute

$$V_{\mathrm{C}} = \mathbb{E}(\hat{\theta}^{\mathrm{C}} - \theta)(\hat{\theta}^{\mathrm{C}} - \theta)^{\mathrm{T}}.$$

The first and second terms are

$$\begin{aligned} &\Pi_{\mathrm{B}} \mathbb{E}(\hat{\theta}^{\mathrm{B}} - \theta)(\hat{\theta}^{\mathrm{B}} - \theta)^{\mathrm{T}} \Pi_{\mathrm{B}}^{\mathrm{T}} + \Pi_{\mathrm{P}} \mathbb{E}(\hat{\theta}^{\mathrm{P}} - \theta)(\hat{\theta}^{\mathrm{P}} - \theta)^{\mathrm{T}} \Pi_{\mathrm{P}}^{\mathrm{T}} \\ &\quad \to [(A+B)^{-1} A \Sigma_{\mathrm{B}} A (A+B)^{-1} + (A+B)^{-1} B \Sigma_{\mathrm{P}} B (A+B)^{-1}] \\ &\quad = (A+B)^{-1} A \frac{1}{\bar{N}} A^{-1} A (A+B)^{-1} + (A+B)^{-1} B \frac{1}{\bar{N}} B^{-1} B (A+B)^{-1} \\ &\quad = \frac{1}{\bar{N}} (A+B)^{-1} \\ &\quad \triangleq \Sigma_{\mathrm{CR}}. \end{aligned}$$

The third and fourth terms are

$$\Pi_{\mathrm{B}} \mathbb{E}(\hat{\theta}^{\mathrm{B}} - \theta)(\hat{\theta}^{\mathrm{P}} - \theta)^{\mathrm{T}} \Pi_{\mathrm{P}}^{\mathrm{T}} + \Pi_{\mathrm{P}} \mathbb{E}(\hat{\theta}^{\mathrm{P}} - \theta)(\hat{\theta}^{\mathrm{B}} - \theta)^{\mathrm{T}} \Pi_{\mathrm{B}}^{\mathrm{T}}.$$

Let us first look at the third term. It can be found that the total number of sample groups $\bar{N}$ is given, and $\hat{\theta}^{\mathrm{B}}$, based on the set-valued measurements, and $\hat{\theta}^{\mathrm{P}}$, based on the precise measurements, are uncorrelated. Therefore, by considering any component of matrix $\mathbb{E}(\hat{\theta}^{\mathrm{B}}-\theta)(\hat{\theta}^{\mathrm{P}}-\theta)^{\mathrm{T}}$, for any $1\leqslant i,j\leqslant n$, we have

$$\mathbb{E}(\hat{\theta}_i^{\mathrm{B}}-\theta_i)(\hat{\theta}_j^{\mathrm{P}}-\theta_j)=\mathbb{E}(\hat{\theta}_i^{\mathrm{B}}-\theta_i)\mathbb{E}(\hat{\theta}_j^{\mathrm{P}}-\theta_j).$$

First, we compute the convergence rate of $\mathbb{E}(\hat{\theta}_i^{\mathrm{B}}-\theta_i)$. Because the estimator $\hat{\theta}^{\mathrm{B}}$ is a linear transformation of $\hat{\psi}^{\mathrm{B}}$, this is equivalent to computing the convergence rate of $\mathbb{E}(\hat{\psi}_i^{\mathrm{B}}-\psi_i)$. We define

$$p_i=\mathbb{E}s_{ij}^{(2)}=F(\psi_i-c),\quad 1\leqslant i\leqslant n, 1\leqslant j\leqslant \bar{N},$$

and a function $T(\cdot)=F^{-1}(x-c)$, which has a second-order continuous derivative. Thus,

$$\mathbb{E}(\hat{\psi}_i^{\mathrm{B}}-\psi_i)=\mathbb{E}\left[T\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{(2)}\right)-T(p_i)\right],$$

where the Taylor expansion in p_i of $T(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{(2)})$ is

$$T'(p_i)\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{(2)}-p_i\right)+\frac{T''(\xi_i)}{2}\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{(2)}-p_i\right)^2,$$

where ξ_i is associated with $\{s_{ij}^{(2)}, 1\leqslant j\leqslant \bar{N}\}$ and p_i. Under Assumption 4, the boundedness of ξ is guaranteed in that $\frac{1}{\bar{N}}\leqslant\xi_i\leqslant\frac{\bar{N}-1}{\bar{N}}$. On account of the continuity of the function $T(\cdot)$, for any $1\leqslant i\leqslant n$, there exists a positive real number M_i so that $T''(\xi_i)\leqslant M_i$. Then

$$\mathbb{E}(\hat{\psi}_i^{\mathrm{B}}-\psi_i)=\mathbb{E}\left[\frac{T''(\xi_i)}{2}\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{1}-p_i\right)^2\right]\leqslant M_i\mathbb{E}\left(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{(2)}-p_i\right)^2.$$

Notice that $\mathbb{E}(\frac{1}{\bar{N}}\sum_{j=1}^{\bar{N}}s_{ij}^{(2)}-p_i)^2=O(\frac{1}{\bar{N}})$; therefore,

$$\mathbb{E}(\hat{\psi}_i^{\mathrm{B}}-\psi_i)=O\left(\frac{1}{\bar{N}}\right),\quad \mathbb{E}(\hat{\theta}_i^{\mathrm{B}}-\theta_i)=O\left(\frac{1}{\bar{N}}\right).$$

Next, we compute the convergence rate of $\mathbb{E}(\hat{\theta}_j^{\mathrm{P}}-\theta_j)$. As before, we need to consider only the convergence rate of $\mathbb{E}(\hat{\psi}_j^{\mathrm{P}}-\psi_j)$. According to the strong law of large numbers, given the asymptotic distribution of $\hat{\psi}_j^{\mathrm{P}}$, it is easy to obtain

$$\mathbb{E}(\hat{\psi}_j^{\mathrm{P}}-\psi_j)=O\left(\frac{1}{\sqrt{\bar{N}}}\right),\quad \mathbb{E}(\hat{\theta}_j^{\mathrm{P}}-\theta_j)=O\left(\frac{1}{\sqrt{\bar{N}}}\right).$$

In this case, we have $\mathbb{E}(\hat{\theta}_i^{\mathrm{B}}-\theta_i)(\hat{\theta}_j^{\mathrm{P}}-\theta_j)=O(\frac{1}{\bar{N}^{3/2}})$. Therefore, the convergence rate of the third term is

$$\Pi_{\mathrm{B}}\mathbb{E}(\hat{\theta}^{\mathrm{B}}-\theta)(\hat{\theta}^{\mathrm{P}}-\theta)^{\mathrm{T}}\Pi_{\mathrm{P}}^{\mathrm{T}}=O\left(\frac{1}{\bar{N}^{3/2}}\right).$$

As for the fourth term, it is the transposed form of the third term. Thus,

$$\Pi_{\mathrm{P}}\mathbb{E}(\hat{\theta}^{\mathrm{P}}-\theta)(\hat{\theta}^{\mathrm{B}}-\theta)^{\mathrm{T}}\Pi_{\mathrm{B}}^{\mathrm{T}}=O\left(\frac{1}{\bar{N}^{3/2}}\right).$$

Because the convergence of the CR lower bound Σ_{CR} is $O(\frac{1}{\bar{N}})$, we can obtain the following:

$$V_{\mathrm{C}}\to\Sigma_{\mathrm{CR}}+O\left(\frac{1}{\bar{N}^{3/2}}\right)\to\Sigma_{\mathrm{CR}},\quad \bar{N}\to\infty.$$

That is,

$$\bar{N}[V_{\mathrm{C}}-\Sigma_{\mathrm{CR}}]\to 0,\quad \bar{N}\to\infty.$$

Notice that $N/\bar{N}=n$; hence, Eq. (28) holds.

Remark 8. It should be emphasized that the two weight matrices Π_{B} and Π_{P} given in Theorem 4 are based on the unknown true system parameter θ. Because of the strong consistency of the two estimators $\hat{\theta}^{\mathrm{B}}$ and $\hat{\theta}^{\mathrm{P}}$ in Theorem 1, Π_{B} and Π_{P} can be constructed by replacing the true parameters with the estimators $\hat{\theta}^{\mathrm{B}}$ and $\hat{\theta}^{\mathrm{P}}$, respectively.

Remark 9. Though the identification problem of the system (2) in a linear form is analyzed, the proposed algorithm can be simply extended to the general system (1). Here we take a more complicated system as an example,

$$\begin{cases} y_k = \phi_k^{\mathrm{T}}\theta + e_k, \\ s_k = \begin{cases} c_2, & \text{if } y_k > c_2, \\ y_k, & \text{if } c_2 \geqslant y_k > c_1, \\ c_1, & \text{otherwise}, \end{cases} \end{cases} \quad k = 1, 2, \ldots, N, \tag{31}$$

where c_1 and c_2 are two set-valued sensor thresholds, and the precise measurements y_k lie in the interval $(c_1, c_2]$. In fact, the above model can be divided into a precise subsystem and two set-valued subsystems in a similar way. The main algorithm process can be described as follows.

(1) Compute three estimators for different subsystems through the ML principle.

(2) Fuse the two estimators for two set-valued subsystems to construct a new estimator. The fusion algorithm can be looked up in [3] and the new estimator has been proved to be consistently unbiased and asymptotically efficient.

(3) Based on the proposed CRBA framework, combine the estimator for precise subsystem with the new estimator for set-valued subsystems to get the final estimation.

5 System identification under general inputs

In this section, we propose an iterative algorithm based on the CRBA for system identification with mixed precise and set-valued measurements under general inputs. As there is no closed-form solution of the ML estimation under general inputs whether for the set-valued or for the precise subsystem, we must adopt iterative algorithms that include the EM algorithm for the set-valued subsystem and the gradient descent algorithm for the precise subsystem.

5.1 The CR lower bound

Following the proof of Theorem 2, we obtain the CR lower bound of the parameter estimation under general inputs; that is,

$$\Sigma_{\mathrm{CR}} = \left(\Phi^{\mathrm{T}}\Sigma_{\phi}^{-1}\Phi\right)^{-1}, \tag{32}$$

where $\Phi = (\phi_1, \phi_2, \ldots, \phi_N)^{\mathrm{T}}$, Σ_{ϕ} is a diagonal matrix and its ith $(1 \leqslant i \leqslant N)$ diagonal element is

$$\Sigma_{\phi}[i,i] = \left[\frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{(1 - F(\phi_i^{\mathrm{T}}\theta - c))F(\phi_i^{\mathrm{T}}\theta - c)} + \frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{F(\phi_i^{\mathrm{T}}\theta - c)} + F(\phi_i^{\mathrm{T}}\theta - c) - f(\phi_i^{\mathrm{T}}\theta - c)(\phi_i^{\mathrm{T}}\theta - c)\right]^{-1}.$$

Let us define two diagonal matrices, $\Sigma_{\mathrm{B}}^{\phi}$ and $\Sigma_{\mathrm{P}}^{\phi}$. Their ith $(1 \leqslant i \leqslant N)$ diagonal elements are

$$\Sigma_{\mathrm{B}}^{\phi}[i,i] = \left[\frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{(1 - F(\phi_i^{\mathrm{T}}\theta - c))F(\phi_i^{\mathrm{T}}\theta - c)}\right]^{-1},$$

$$\Sigma_{\mathrm{P}}^{\phi}[i,i] = \left[\frac{f^2(\phi_i^{\mathrm{T}}\theta - c)}{F(\phi_i^{\mathrm{T}}\theta - c)} + F(\phi_i^{\mathrm{T}}\theta - c) - f(\phi_i^{\mathrm{T}}\theta - c)(\phi_i^{\mathrm{T}}\theta - c)\right]^{-1}.$$

Then, the CR lower bound Σ_{CR} can be written as

$$\Sigma_{\mathrm{CR}} = \left(\Phi^{\mathrm{T}}[\Sigma_{\mathrm{B}}^{\phi}]^{-1}\Phi + \Phi^{\mathrm{T}}[\Sigma_{\mathrm{P}}^{\phi}]^{-1}\Phi\right)^{-1}. \tag{33}$$

It can be inferred that the former part $\Phi^{\mathrm{T}}[\Sigma_{\mathrm{B}}^{\phi}]^{-1}\Phi$, and the latter part $\Phi^{\mathrm{T}}[\Sigma_{\mathrm{P}}^{\phi}]^{-1}\Phi$ in (33) are associated with the covariances of the estimators for the set-valued subsystem and for the precise subsystem, respectively. Because of the special structure of the CR lower bound, the CRBA algorithm can be extended to the general situation.

5.2 Gradient descent algorithm for the precise subsystem

In a similar approach to that used in Subsection 3.2.1, let us assume that the former N_1 measurements $\{s_i^{(2)}, 1 \leqslant i \leqslant N_1\}$ are precise data. The log-likelihood function for the overall probability of the measurements, given the inputs Φ_N and the parameter θ, is

$$l_{\mathrm{p}}(\theta) = \sum_{i=1}^{\bar{N}_1} \log \frac{f(s_{ij}^2 - \phi_i^{\mathrm{T}}\theta)}{1 - F(c - \phi_i^{\mathrm{T}}\theta)}. \tag{34}$$

A common optimization algorithm—the gradient descent algorithm is chosen as the identification method for the precise subsystem under general inputs. The iterative formula is as follows:

$$\hat{\theta}_{t+1}^{\mathrm{P}} = \hat{\theta}_t^{\mathrm{P}} + \lambda_t p_t, \tag{35}$$

where p_t is the direction of search, namely, the negative gradient $p_t = - \bigtriangledown l_{\mathrm{p}}(\hat{\theta}_t^{\mathrm{B}})$ and λ_t is step size.

5.3 EM algorithm for the set-valued subsystem

Similar to what was shown in Subsection 3.2.2, the log-likelihood function for the overall probability of the measurements $\{s_i^1, 1 \leqslant i \leqslant N\}$, given the inputs Φ_N and the parameter θ, is as follows:

$$l_{\mathrm{b}}(\theta) = \sum_{i=1}^{N} \log[1 - F(c - \phi_i^{\mathrm{T}}\theta)]\mathbb{I}_{\{s_i^1=1\}} + \log[F(c - \phi_i^{\mathrm{T}}\theta)]\mathbb{I}_{\{s_i^1=0\}}. \tag{36}$$

With reference to the work of Zhao et al. [25], the EM algorithm is chosen as the identification method for the set-valued subsystem under general inputs, because it has been proved to converge to the ML estimation with an exponential rate. The iterative formula is as follows:

$$\hat{\theta}_{t+1}^{\mathrm{B}} = \hat{\theta}_t^{\mathrm{B}} - \left(\sum_{k=1}^{N} \phi_k \phi_k^{\mathrm{T}}\right)^{-1} \left(\sum_{k=1}^{N} \phi_k \cdot f(c - \phi_k^{\mathrm{T}}\hat{\theta}_t^{\mathrm{B}}) \left[\frac{\mathbb{I}_{[s_k=0]}}{F(c - \phi_k^{\mathrm{T}}\hat{\theta}_t^{\mathrm{B}})} - \frac{\mathbb{I}_{[s_k=1]}}{1 - F(c - \phi_k^{\mathrm{T}}\hat{\theta}_t^{\mathrm{B}})}\right]\right), \tag{37}$$

where t is the iterative time of estimation.

5.4 CRBA for general inputs

Similar to Theorem 3, two weight matrices are defined based on the above CR lower bounds:

$$\Pi_{\mathrm{B}} = (\Sigma_{\mathrm{B}}^{-1} + \Sigma_{\mathrm{P}}^{-1})^{-1}\Sigma_{\mathrm{B}}^{-1}, \tag{38}$$

$$\Pi_{\mathrm{P}} = (\Sigma_{\mathrm{B}}^{-1} + \Sigma_{\mathrm{P}}^{-1})^{-1}\Sigma_{\mathrm{P}}^{-1}, \tag{39}$$

where $\Sigma_{\mathrm{B}} = \Phi^{-1}\Sigma_{\mathrm{B}}^{\phi}\Phi^{-\mathrm{T}}$ and $\Sigma_{\mathrm{P}} = \Phi^{-1}\Sigma_{\mathrm{P}}^{\phi}\Phi^{-\mathrm{T}}$. Finally, the CRBA is constructed as

$$\hat{\theta}^{\mathrm{C}} = \Pi_{\mathrm{B}}\hat{\theta}^{\mathrm{B}} + \Pi_{\mathrm{P}}\hat{\theta}^{\mathrm{P}}. \tag{40}$$

Because the proposed algorithm consists of two iterative algorithms, the iterative form of the CRBA can be designed as follows:

$$\hat{\theta}_t^{\mathrm{C}} = \Pi_{\mathrm{B}}(\hat{\theta}_{t-1}^{\mathrm{C}})\hat{\theta}_t^{\mathrm{B}} + \Pi_{\mathrm{P}}(\hat{\theta}_{t-1}^{\mathrm{C}})\hat{\theta}_t^{\mathrm{P}}, \tag{41}$$

where the initial value $\hat{\theta}_0^{\mathrm{C}}$ can be chosen from the estimations $\hat{\theta}_1^{\mathrm{B}}$ and $\hat{\theta}_1^{\mathrm{P}}$.

6 The difference between the CRBA and the JMLE

In this section, the difference between the CRBA and the JMLE is discussed in terms of the principle of optimality and variance analysis.

6.1 The principle of optimality

The CRBA and JMLE methods can be described as

$$\hat{\theta}^{\mathrm{J}} = \arg\max_{\theta} l(\theta) = \arg\max_{\theta}\{l_{\mathrm{b}}(\theta) + l_{\mathrm{p}}(\theta)\},$$
$$\hat{\theta}^{\mathrm{C}} = \Pi_{\mathrm{B}}\hat{\theta}^{\mathrm{B}} + \Pi_{\mathrm{P}}\hat{\theta}^{\mathrm{P}} = \Pi_{\mathrm{B}}\arg\max_{\theta} l_{\mathrm{b}}(\theta) + \Pi_{\mathrm{P}}\arg\max_{\theta} l_{\mathrm{p}}(\theta),$$

where $l(\theta)$, $l_{\mathrm{b}}(\theta)$, and $l_{\mathrm{p}}(\theta)$ are the log-likelihood functions based on the whole system, the binary-valued subsystem, and the precise subsystem, respectively.

To explain the difference between the two methods clearly, a single parameter system is proposed as an example, where the weights Π_{B} and Π_{P} are scalars and satisfy $\Pi_{\mathrm{B}} + \Pi_{\mathrm{P}} = 1$. Therefore, $\hat{\theta}^{\mathrm{J}}$ can be written as

$$\begin{aligned}\hat{\theta}^{\mathrm{J}} &= \arg\max_{\theta}\{0.5(l_{\mathrm{b}}(\theta) + l_{\mathrm{p}}(\theta))\}\\ &= \arg\max_{\theta}\{0.5l_{\mathrm{b}}(\theta) + 0.5l_{\mathrm{p}}(\theta)\}.\end{aligned}$$

In short, the CRBA corresponds to a problem concerning the weights of different estimators, whereas the JMLE corresponds to a problem concerning the weights of different log-likelihood functions. To put this differently, the optimal CRBA is obtained by finding the weights that make the variance of the CRBA smallest, whereas the JMLE gives the same weight to two log-likelihood functions that have asymptotic variances larger than that of the optimal CRBA for a system with periodic inputs. As for the system with general inputs, we propose a special CRBA based on the CR lower bounds of two ML estimations for binary-valued and precise subsystems, although their optimality cannot be guaranteed. Fortunately, extensive numerical simulations have demonstrated the excellent performance of the special CRBA. These results raise the prospect that whether one can obtain a better estimator than the JMLE and even of achieving the CR lower bound, as well as the optimal CRBA, by choosing appropriate weights for the two log-likelihood functions.

6.2 Variance analysis

To illustrate the difference between the CRBA and the JMLE in terms of their asymptotic efficiency, we will take as an example a one-dimensional system under periodic inputs. The threshold $c = 0$ and the system input is a constant 1.

We can generate different proportions of the precise information by changing the system parameter θ. Figure 1 shows the difference in the asymptotic variance and the CR lower bound for different identification methods including the JMLE based on the whole system and the MLE based on two subsystems. The longitudinal axis represents the product of the sample size and the difference in the asymptotic variance and the CR lower bound.

As it has been proved above that the CRBA can achieve the CR lower bound, we will focus on the variances based on two-subsystem identification. With an increasing rate of precise information, the variance of the set-valued subsystem identification increases gradually, but the variances of the precise subsystem identification and the JMLE both decrease gradually. As expected, the variance of the JMLE is much smaller than that of the precise subsystem identification because of a lack of samples.

There may exist some confusion about that the JMLE cannot asymptotically achieve the CR lower bound though the theoretical analysis is given in Theorem 3. Generally speaking, the ML estimator can asymptotically achieve the CR lower bound. But the JMLE is not an ML estimator in a strict sense and is just a pseudo-ML estimator for the identification problem. From the deductive process, Eq. (11) is actually the ML equation and cannot be solved by giving an exact solution. And then on the basis of (14), the ML estimator can be approximately expressed as the JMLE. Furthermore, the JMLE is just the moment estimator from (12) rather than the ML estimator. So this can explain why JMLE cannot asymptotically achieve the CR lower bound.

Furthermore, when the rate of precise information is relatively smaller (e.g., less than 0.4 in Figure 1), the variance of the set-valued subsystem identification is closer to the CR lower bound than the variances of the JMLE and the precise subsystem identification. However, as the rate of precise information increases, the variance of the JMLE is closer to the CR lower bound than the variance of the two-subsystem identification and, to a certain degree (e.g., more than 0.8 in Figure 1), the precise subsystem

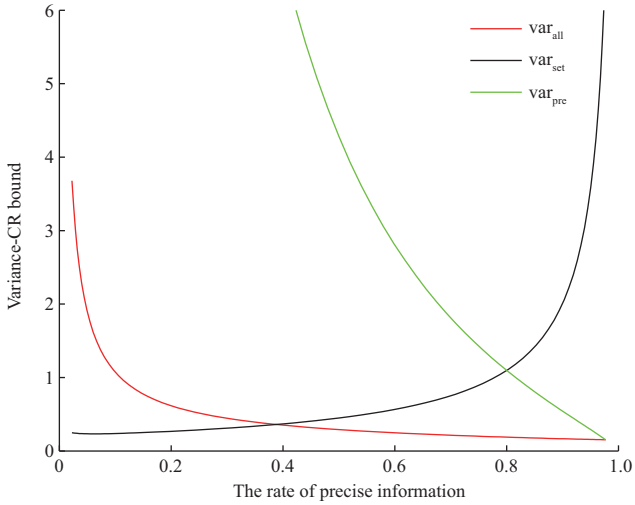

Figure 1 (Color online) Variance comparison for different proportions of precise data.

identification performs better than the set-valued subsystem identification. Above all, it is obvious that the CRBA performs better only in cases when there is less precise information.

As for general inputs, we cannot analyze the variance of the different identification methods in detail theoretically, because the actual estimators may not be available. Fortunately, it can be shown via numerical simulations that the CRBA consisting of the EM algorithm and the gradient descent algorithm maintains its advantage.

7 Numerical simulations

In this section, we conduct extensive simulations to illustrate the main theoretical results in different situations and demonstrate the identification effects of the CRBA in terms of consistency and asymptotic efficiency.

7.1 Systems with periodic inputs

Consider a three-dimensional system with the parameter vector $\theta = (-1, 1, 3)^{\mathrm{T}}$ and threshold $c = 0$. The period of system input is set to be the same as the dimension of the system. We then generate the precise information for different proportions by changing the periodic inputs in the model (1). Two different situations are considered.

(A1) For each periodic input $\phi_i, i = 1, 2, 3$, the corresponding precise data s_i is relatively low (less than half of total information obtained). Here, the three periodic inputs are set as

$$\phi_1 = (1, 1, 0)^{\mathrm{T}},\ \phi_2 = (0.5, -2, 0.5)^{\mathrm{T}},\ \phi_3 = (0, -1, -1)^{\mathrm{T}}.$$

(A2) For each periodic input $\phi_i, i = 1, 2, 3$, the corresponding precise data s_i is moderate or relatively high (representing the majority of the total information obtained). Here, the three periodic inputs are set as

$$\phi_1 = (1, -1.5, 1)^{\mathrm{T}},\ \phi_2 = (2, 2, 0)^{\mathrm{T}},\ \phi_3 = (3, -0.5, 1)^{\mathrm{T}}.$$

The related simulation results with JMLE($\hat{\theta}^{\mathrm{J}}$) and CRBA ($\hat{\theta}^{\mathrm{C}}$) are shown in Figures 2 and 3. The left panels in Figures 2 and 3 show the parameter estimation errors for $\hat{\theta}^{\mathrm{C}}$ and $\hat{\theta}^{\mathrm{J}}$, respectively. The right panels show the empirical variances of $\hat{\theta}^{\mathrm{C}}$ and $\hat{\theta}^{\mathrm{J}}$, respectively. Finally, it can be concluded that as the sample size increases, both estimators converge to the true value and the empirical variance of $\hat{\theta}^{\mathrm{C}}$ becomes closer to the CR lower bound than that of $\hat{\theta}^{\mathrm{J}}$ when precise data is small but the superiority does not appear obvious when the proportion of precise data is greater.

In summary, the estimator $\hat{\theta}^{\mathrm{C}}$, as compared to $\hat{\theta}^{\mathrm{J}}$, can achieve the CR lower bound in any case, thus demonstrating a better identification effect.

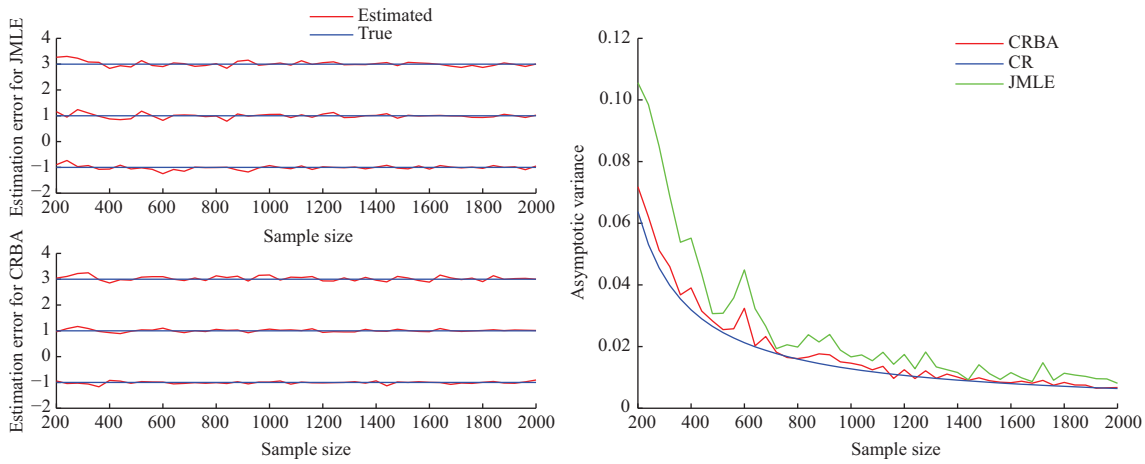

Figure 2 (Color online) Results with a low proportion of precise data under periodic inputs with 100 replicated simulations.

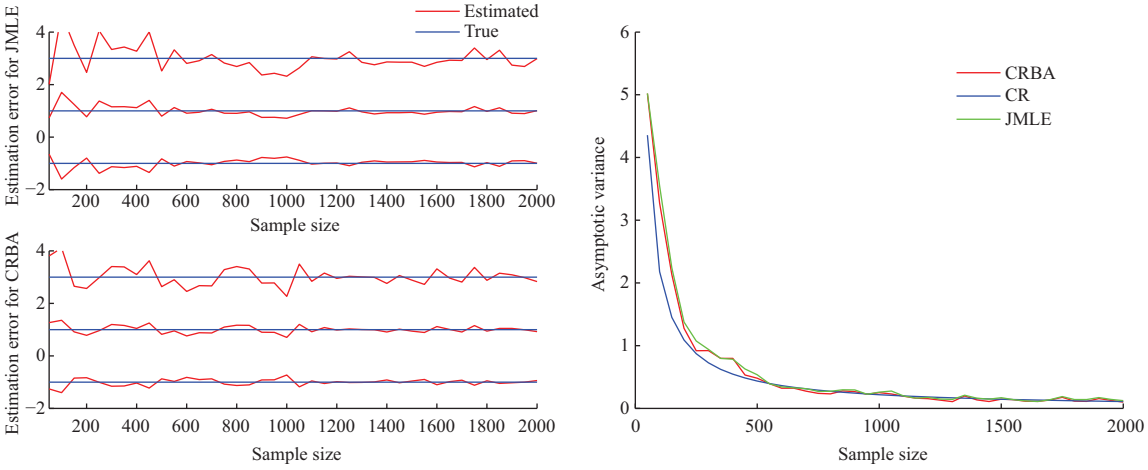

Figure 3 (Color online) Results with a moderate or high proportion of precise data under periodic inputs with 100 replicated simulations.

7.2 Systems with general inputs

For general input, we will take as an example a one-dimensional system with the parameter $(\theta, c) = (2, 0)$. As before, we will consider two situations with different proportions of precise information by changing the general inputs. For comparison, for the precise subsystem, we will adopt the gradient descent algorithm as the identification method for the JMLE.

(B1) The proportion of precise information is relatively low. Here, the inputs Φ are generated based on a uniform distribution (MATLAB code as below).

$$\Phi = -1 + 0.2 \times \mathrm{rand}(N, 1).$$

(B2) The proportion of precise information is relatively high or moderate. Here, the inputs Φ are generated based on a uniform distribution (MATLAB code as below).

$$\Phi = 0.2 - 0.2 \times \mathrm{rand}(N, 1).$$

For a one-dimensional system with general inputs, the related simulation results of the JMLE ($\hat{\theta}^{\mathrm{J}}$) and CRBA ($\hat{\theta}^{\mathrm{C}}$) can be seen in Figures 4 and 5. Although we cannot obtain their closed-form solutions for the MLE, we can conclude that the CRBA based on the EM algorithm and the gradient descent algorithm has superior asymptotic efficiency.

The above simulations show the relative performance of the proposed algorithm CRBA compared with a baseline JMLE. We can see that two algorithms both show good performance when the proportion of

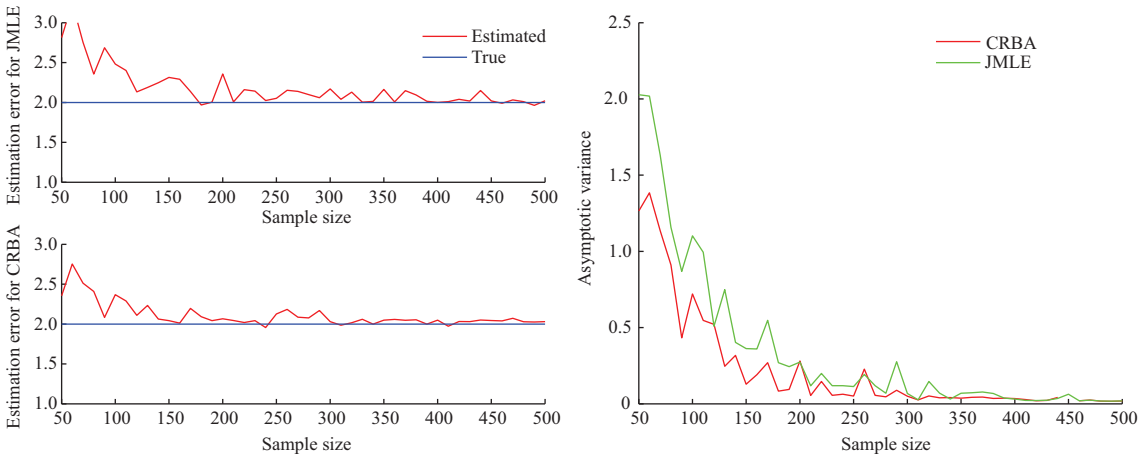

Figure 4 (Color online) Results with a low proportion of precise data under general inputs with 500 replicated simulations.

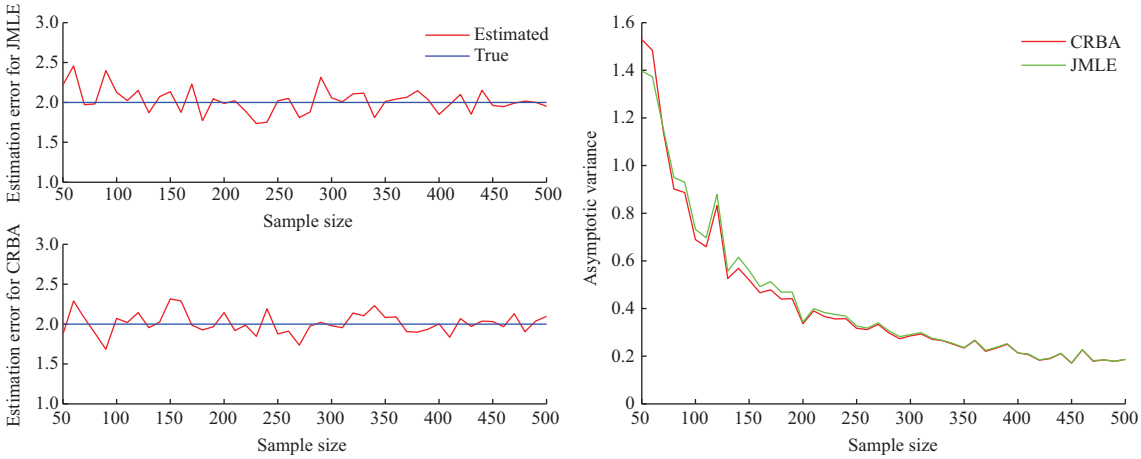

Figure 5 (Color online) Results with a greater proportion of precise data under general inputs with 500 replicated simulations.

precise data is high, but a low proportion has a less serious effect on CRBA than JMLE. In order to further figure out the impact of the proportion of the precise data on the performance of the proposed algorithm, some simulations are conducted to analyze the relation between the ratio of asymptotic variance and CR lower bound and the proportion of precise data under periodic or general inputs.

Figure 6 shows the ratio of asymptotic variance and CR lower bound when the proportion of precise data varies from low to high. Even though the proportion is low, the asymptotic variance is still much close to the CR lower bound. It can be concluded from the results that the proportion of the precise data almost has little effect on the proposed algorithm, which proves the efficiency and stability of the proposed algorithm.

8 Summary

In this paper, we have investigated an identification problem concerning systems with mixed precise and set-valued measurements. The fact that the CR lower bound can be divided into two independent parts prompted us to adopt the concept of first classification and then fusion. That is, we first select the corresponding identification method for each type of data separately and then combine both into the optimal estimation by adding suitable weights, which are chosen based on the variance of their respective estimations.

We have proposed a CR-based algorithm that combines the separate estimations of the precise and set-valued data. As a comparison, the ML estimation based on the JMLE was also constructed. The convergence properties of the algorithms were analyzed in terms of their consistency and asymptotic efficiency under periodic inputs. The CRBA, as opposed to the JMLE, was proved to achieve the CR

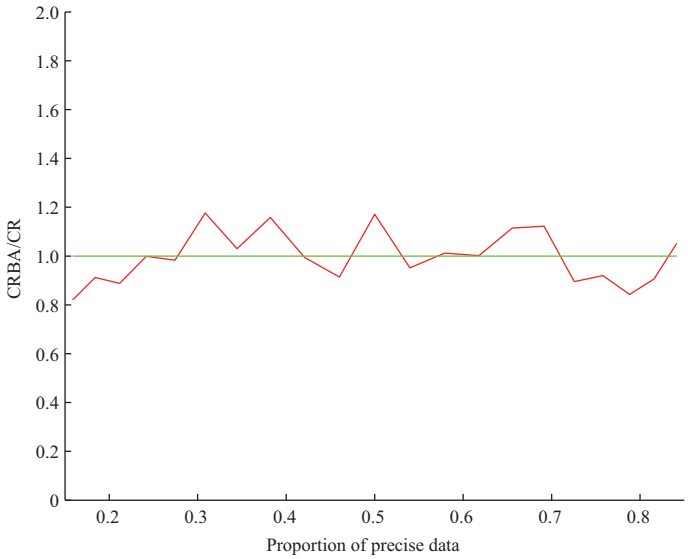

Figure 6 (Color online) The ratio of asymptotic variance and CR lower bound and the proportion of precise data.

lower bound asymptotically. Furthermore, an iterative algorithm based on the CRBA that combines the EM algorithm for the binary-valued subsystem and the gradient descent algorithm for the precise subsystem under general inputs has been constructed, and numerical simulations have shown that its identification effect is superior to that of the JMLE method based on the gradient descent algorithm.

Within the framework outlined in this paper, there are still many meaningful problems that merit further research. For example, how should we handle identification problems when the set-valued measurements are multi-valued, the thresholds and distributions of system noise are unknown, and/or the system structures turn out to be more complicated?

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2018YFA0703800), National Natural Science Foundation of China (Grant No. 62025306), and CAS Project for Young Scientists in Basic Research (Grant No. YSBR-008).

References

1 Ljung L. Perspectives on system identification. Annu Rev Control, 2010, 34: 1–12
2 Ljung L. System Identification: Theory for the User. 2nd ed. Englewood Cliffs: Prentice-Hall, 1999
3 Wang L Y, Yin G G, Zhang J F, et al. System Identification With Quantized Observations. Boston: Birkhäuser, 2010
4 Liu Y, Zhao J, Xia M Z, et al. Model reference adaptive control-based speed control of brushless DC motors with low-resolution hall-effect sensors. IEEE Trans Power Electron, 2014, 29: 1514–1522
5 Ramsden E. Hall-Effect Sensors: Theory and Application. Burlington: Elsevier, 2006
6 Tan S P, Guo J, Zhao Y L, et al. Adaptive control with saturation-constrainted observations for drag-free satellites—a set-valued identification approach. Sci China Inf Sci, 2021, 64: 202202
7 Guo S J, Zeng D L. An overview of semiparametric models in survival analysis. J Stat Plan Infer, 2014, 151–152: 1–16
8 He P, Kong G, Su Z. Estimating the survival functions for right-censored and interval-censored data with piecewise constant hazard functions. Contemporary Clin Trials, 2013, 35: 122–127
9 Boyd S D, Marshall E L, Merker J D, et al. Measurement and clinical monitoring of human Lymphocyte clonality by massively parallel V-D-J pyrosequencing. Sci Transl Med, 2009, 1: 12–23
10 O'Connor D, Enshaei A, Bartram J, et al. Genotype-specific minimal residual disease interpretation improves stratification in pediatric acute lymphoblastic leukemia. J Clin Oncol, 2018, 36: 34–43
11 Sullivan E M, Jeha S, Kang G L, et al. NK cell genotype and phenotype at diagnosis of acute lymphoblastic leukemia correlate with postinduction residual disease. Clin Cancer Res, 2014, 20: 5986–5994
12 Tobin J. Estimation of relationships for limited dependent variables. Econometrica, 1958, 26: 24–36
13 Åström K J. Maximum likelihood and prediction error methods. Automatica, 1980, 16: 551–574
14 Clarke D W. Generalized least squares estimation of the parameters of a dynamic model. In: Proceedings of the IFAC Symposium on Identification of Automatic Control Systems, 1967. 1–11
15 Chen T S, Ohlsson H, Ljung L. On the estimation of transfer functions, regularizations and Gaussian processes-revisited. Automatica, 2012, 48: 1525–1535
16 Pillonetto G, Dinuzzo F, Chen T, et al. Kernel methods in system identification, machine learning and function estimation: a survey. Automatica, 2014, 50: 657–682
17 Söderström T. Errors-in-variables methods in system identification. Automatica, 2007, 43: 939–958
18 Söderström T, Soverini U. Errors-in-variables identification using maximum likelihood estimation in the frequency domain. Automatica, 2017, 79: 131–143
19 Yu C P, Zhang C S, Xie L H. Blind system identification using precise and quantized observations. Automatica, 2013, 49: 2822–2830
20 Wang L Y, Zhang J-F, Yin G G. System identification using binary sensors. IEEE Trans Automat Contr, 2003, 48: 1892–1907

21 Wang T, Tan J W, Zhao Y L. Asymptotically efficient non-truncated identification for FIR systems with binary-valued outputs. Sci China Inf Sci, 2018, 61: 129208
22 Zhang H, Wang T, Zhao Y L. FIR system identification with set-valued and precise observations from multiple sensors. Sci China Inf Sci, 2019, 62: 052203
23 Marelli D, You K, Fu M. Identification of ARMA models using intermittent and quantized output observations. Automatica, 2013, 49: 360–369
24 Godoy B I, Goodwin G C, Agüero J C, et al. On identification of FIR systems having quantized output data. Automatica, 2011, 47: 1905–1915
25 Bi W J, Zhao Y L. Iterative parameter estimate with batched binary-valued observations: convergence with an exponential rate. IFAC Proc Volumes, 2014, 47: 3220–3225
26 Guo J, Zhao Y L. Recursive projection algorithm on FIR system identification with binary-valued observations. Automatica, 2013, 49: 3396–3401
27 Wang T, Zhang H, Zhao Y L. Parameter estimation based on set-valued signals: theory and application. Acta Math Appl Sin Engl Ser, 2019, 35: 255–263
28 Li X Q, Xu Z G, Cui J R, et al. Suboptimal adaptive tracking control for FIR systems with binary-valued observations. Sci China Inf Sci, 2021, 64: 172202
29 Wang Y, Zhao Y, Zhang J F. Distributed recursive projection identification with binary-valued observations. J Syst Sci Complex, 2021, 34: 2048–2068
30 Jing L, Zhang J F. LS-based parameter estimation of DARMA systems with uniformly quantized observations. J Syst Sci Complex, 2022, 35: 748–765
31 Xiao J, Song Q. Recursive identification of quantized linear systems. J Syst Sci Complex, 2019, 32: 985–996
32 Fair R C. A note on the computation of the tobit estimator. Econometrica, 1977, 45: 1723–1727
33 Heckman J J. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator of such models. Ann Econ Soc Meas, 1976, 5: 475–492
34 Wang L Y, Yin G G, Zhang J F. Joint identification of plant rational models and noise distribution functions using binary-valued observations. Automatica, 2006, 42: 535–547
35 Lehmann E L. Elements of Large-Sample Theory. Berlin: Springer, 1998