Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Robust Interval Prediction Model Identification with A Posteriori Reliability Guarantee ⋆

Chao Wang * **Chao Shang** * **Fan Yang** * **Dexian Huang** *
Bin Yu **

* *Department of Automation, Tsinghua University, and Beijing National Research Center for Information Science and Technology, Beijing 100084, China (e-mail: chao-wan17@mails.tsinghua.edu.cn, c-shang@tsinghua.edu.cn, yangfan@tsinghua.edu.cn, huangdx@tsinghua.edu.cn)*
** *Hengli Petrochemical Co., Ltd., Dalian 116001, China (e-mail: yubin@hengli.com)*

Abstract: In classical paradigm of model identification, a single prediction value is returned as a point estimate of the output. Recently, the interval prediction model (IPM) has been receiving increasing attentions. Different from generic models, an IPM gives an interval of confidence as the prediction that covers the majority of training data while being as tight as possible. However, due to the randomness of sampling training data, the reliability of IPM constructed is uncertain. In this paper, we focus on a general class of IPMs where a fraction of data samples can be discarded to pursue robustness, and establish an appropriate a posteriori reliability guarantee. It relies on counting the "decisive" constraints associated with the optimal solution, and generally leads to reduced conservatism and better estimation performance than the existing performance bounds. Moreover, the guarantee holds irrespective of the data generation mechanism, which informs the decision maker of the prediction confidence in the absence of precise knowledge about data distribution. Its effectiveness is illustrated based on numerical examples.

Keywords: Interval prediction model, system identification, uncertainty quantification, robustness.

1. INTRODUCTION

In standard routes of system identification, one typically uses a parametric model to fit the observed input-output data, and then makes predictions on out-of-sample inquiry inputs. The prediction value, however, will be no longer informative if its accuracy is insufficient (Box et al. (2015)). Hence it is important to not only deliver the prediction value but also establish a confidence interval around the prediction value as an indicator of reliability and accuracy.

For effective uncertainty quantification, interval predictor model (IPM), which returns an interval instead of a point estimate as the output, has been receiving consistent attentions since its invention. Early contributions in this vein include Chryssolouris et al. (1996); Heskes (1997); Jaulin et al. (2001); Calafiore and Campi (2002), in different forms of interval analysis, set-membership prediction, confidence interval prediction, and so on. Later on, widespread applications have been motivated in multifarious areas, such as energy clearing price prediction (Zhang et al. (2003)), river flow prediction (Shrestha and Solomatine (2006)), demand prediction in power systems (Morita et al. (1996)), sensor fault detection (Fravolini et al. (2018)), model predictive control (Shang and You (2019a)), and chemical process planning (Shang et al. (2017)).

For traditional parametric models, after the structure of the system (e.g. some basis functions) is chosen and model parameters are determined, the confidence interval is then constructed, which calls for the statistical postulation made upon the true system, e.g. Gaussian-distributed measurement noise (Ljung (1999)). If data distribution is well characterized by statistical assumptions, one can effectively recover the system and achieve desirable prediction performance. However, in some cases of engineering practice, the exact probabilistic distribution based on which data are generated is typically unknown. If inconsistency exists between the physical truth and the underlying assumption, then a heavily biased estimation of confidence interval and model reliability will be attained.

To address this issue, we establish an *a posteriori distribution-free* guarantee for reliability of induced intervals for a class of robust IPMs, where the outcome interval covers the majority of observed outputs with the interval being as "narrow" as possible to ensure accuracy. Specifically, a fraction of data samples are allowed to reside outside the interval, which leads to reduced conservatism but enhanced robustness against outliers (Campi et al. (2009)). By learning from data and excavating structural

⋆ This work was supported by the National Natural Science Foundation of China (Nos. 61673236, 61433001, 61873142), and the Industrial Internet Innovative Development Project in 2019 (No. TC19084Y).

information within model parameters, which is expressed in terms of the number of "decisive" data samples, we show that the reliability, i.e. the probability that a new sample falls into the interval, can be bounded probabilistically and in a solution-dependent way. Most importantly, the reliability guarantee exhibits a distribution-free nature, implying that no further assumptions on the data-generating mechanism are needed. In comparison with existing a priori bounds in literature, the proposed a posteriori bound has flexibility to adapt to the quality of realized model parameters on training samples, and shows significantly reduced conservatism in estimation performance. These merits will be illustrated by numerical case studies.

2. ROBUST INTERVAL PREDICTOR MODEL

2.1 Problem setup

First we clarify some notations and definitions. We define a set of continuous integer $\mathbb{N}_{a:b} = \{a, a+1, \cdots, b-1, b\}$. $C^u[0,1]$ denotes a class of u times differentiable functions with continuous u-th order derivative over $[0,1]$. $\mathbf{1}_A(\cdot)$ is the indicator function over set A. For two symmetric matrices X and Y, $X \succeq Y$ denotes positive semi-definiteness of $X - Y$.

We consider a system with $\boldsymbol{\phi} \in \Phi \subseteq \mathbb{R}^m$ as the input and $y \in Y \subseteq \mathbb{R}$ as the output, where Φ and Y are the input set and the output set, respectively. It is assumed that the mechanism of generating data samples $\{\boldsymbol{\phi}_i, y_i\}_{i=1}^N$ is stationary, which is denoted with $\mathbb{P}$ and can be interpreted as a joint probability measure in the $\Phi \times Y$ space (Campi et al. (2009)). Although $\mathbb{P}$ is unknown, a set of input-output data $\{(\boldsymbol{\phi}_i, y_i) | i \in \mathbb{N}_{1:N}\}$ can be randomly sampled, which are assumed to be independent from each other. Such requirements of stationarity and independence just indicate that the samples $\{(\boldsymbol{\phi}_i, y_i) | i \in \mathbb{N}_{1:N}\}$ are independently and identically distributed (i.i.d).

Note that along generic routes of system identification, the ideal output y is regarded as a function from $\boldsymbol{\phi}$ to y: $y = f(\boldsymbol{\phi})$. Rather, in our scheme the induced y could be different with the same input $\boldsymbol{\phi}$ due to various sources of uncertainty. In this case, predicting the possible range of output with an interval makes more sense than giving a single value.

Formally speaking, an IPM gives a suitable output interval in Y for each possible input $\boldsymbol{\phi}$, which can be seen as a *set-valued mapping* $F(\boldsymbol{\phi})$ in $\boldsymbol{\phi}$, i.e., $\boldsymbol{\phi} \to F(\boldsymbol{\phi}) \subseteq Y$. As the true mechanism $\mathbb{P}$ is unknown, the goal of IPM identification is to use a series of i.i.d samples $\{(\boldsymbol{\phi}_i, y_i) | i \in \mathbb{N}_{1:N}\}$ obtained from the system to find the mapping $F(\boldsymbol{\phi})$. It is hoped that the interval can describe the distribution of output y while being adequately informative. In other words, the interval shall contain the majority of possible outputs with a small volume. This inspires the following general formulation of IPM identification from N i.i.d. samples:

$$\begin{aligned} \text{IPM}_N: \quad & \min_{F} \ \mathbb{E}_{\boldsymbol{\phi}} \{\text{Size of } F(\boldsymbol{\phi})\} \\ & \text{s.t.} \ \ y_i \in F(\boldsymbol{\phi}_i), \ \forall i \in \mathbb{N}_{1:N}. \end{aligned} \tag{1}$$

To construct the set-valued mapping $F(\boldsymbol{\phi})$, one first chooses a proper parametric function $\xi = g(\boldsymbol{\phi}, \boldsymbol{\eta})$, which is a single-valued mapping from Φ to Y with $\boldsymbol{\eta}$ being model parameters. For a fixed $\boldsymbol{\phi}$, $F(\boldsymbol{\phi})$ can be constructed by considering possible values of ξ induced by variations of $\boldsymbol{\eta}$ in a set Γ (Campi et al. (2009)):

$$F(\boldsymbol{\phi}) = \{\xi | \xi = g(\boldsymbol{\phi}, \boldsymbol{\eta}), \forall \boldsymbol{\eta} \in \Gamma\}. \tag{2}$$

Note that the function $\xi = g(\boldsymbol{\phi}, \boldsymbol{\eta})$ used here has to be continuous. Under this circumstance, if $\boldsymbol{\eta}$ is distributed in a convex set, for any given $\boldsymbol{\phi}$, the value of $\xi = g(\boldsymbol{\phi}, \boldsymbol{\eta})$ will be distributed in an interval.

Obviously, the single-valued function $g(\boldsymbol{\phi}, \boldsymbol{\eta})$ and the set Γ for $\boldsymbol{\eta}$ specify the set-valued mapping $F(\boldsymbol{\phi})$. Once we have chosen the parametric form of $g(\boldsymbol{\phi}, \boldsymbol{\eta})$, the optimization problem IPM_N is a model with Γ as decision variables. In this paper, the simple affine mapping $g(\boldsymbol{\phi}, \boldsymbol{\eta})$ is considered:

$$\xi = g(\boldsymbol{\phi}, \boldsymbol{\eta}) = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\phi} + e, \tag{3}$$

where $\boldsymbol{\eta} := \{\boldsymbol{\theta}, e\}$ denotes model parameters.

Remark 1. Note that the input $\boldsymbol{\phi}$ does not have to coincide with real inputs of the system, and a linear structure has abundant modeling power. In fact, it can be an appropriate mapping of real inputs after transformations to attain a nonlinear IPM. For example, if x is a one-dimensional real input of the system, then $\boldsymbol{\phi}$ can be chosen as $\boldsymbol{\phi} = [x, x^2, x^3, \dots]$ as a polynomial expansion. Other nonlinear basis functions such as wavelets can be adopted as well.

Two sets are individually designed to describe possible variations in $\boldsymbol{\theta}$ and e. For the set of $\boldsymbol{\theta}$, a standard choice is to use a ball or an ellipsoid with center $\boldsymbol{c}$:

$$\{\boldsymbol{\theta} \mid ||\boldsymbol{\theta} - \boldsymbol{c}|| \le r\}, \tag{4}$$

$$\{\boldsymbol{\theta} \mid (\boldsymbol{\theta} - \boldsymbol{c})^{\mathrm{T}} P^{-1} (\boldsymbol{\theta} - \boldsymbol{c}) \le 1\}, \tag{5}$$

while the set of e can be chosen as a simple symmetric interval $[-\gamma, \gamma]$. Under such choices, the induced set-valued mapping $F(\boldsymbol{\phi})$ can be expressed as (Calafiore and Campi (2003)):

$$\left[\boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi} - (r||\boldsymbol{\phi}|| + \gamma), \boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi} + (r||\boldsymbol{\phi}|| + \gamma)\right], \tag{6}$$

$$\left[\boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi} - \left(\sqrt{\boldsymbol{\phi}^{\mathrm{T}} P \boldsymbol{\phi}} + \gamma\right), \boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi} + \left(\sqrt{\boldsymbol{\phi}^{\mathrm{T}} P \boldsymbol{\phi}} + \gamma\right)\right]. \tag{7}$$

In these two cases, the identification of corresponding IPMs can be described as tractable optimization problems (Calafiore and Campi (2003)). For IPM (6) it suffices to solve a linear programming program:

$$\min_{\boldsymbol{c}, r, \gamma} \ \alpha r + \gamma \tag{8a}$$

$$\text{s.t.} \ \ -r||\boldsymbol{\phi}_i|| - \gamma \le y_i - \boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi}_i \le r||\boldsymbol{\phi}_i|| + \gamma, \forall i, \tag{8b}$$

$$r \ge 0, \gamma \ge 0, \tag{8c}$$

where $\alpha > 0$ is a weighting parameter to balance between sizes of two sets. For IPM (7) induced by the ellipsoidal parameter set, the following semi-definite program (SDP) shall be resolved (Campi et al. (2009)):

$$\min_{\boldsymbol{c}, P, \gamma^2, \epsilon_1, \dots, \epsilon_N} \ \mathrm{Tr}[PW] + \gamma^2 \tag{9a}$$

$$\text{s.t.} \quad P \succ 0, \ \begin{bmatrix} \gamma^2 & \epsilon_i \\ \epsilon_i & 1 \end{bmatrix} \succeq 0, \ \forall i \in \mathbb{N}_{1:N}, \tag{9b}$$

$$\begin{bmatrix} \boldsymbol{\phi}_i^{\mathrm{T}} P \boldsymbol{\phi}_i & y_i - \boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi}_i - \epsilon_i \\ y_i - \boldsymbol{c}^{\mathrm{T}} \boldsymbol{\phi}_i - \epsilon_i & 1 \end{bmatrix} \succeq 0, \tag{9c}$$

where W is a weighting matrix similar to α, and $\{\epsilon_i, i \in \mathbb{N}_{1:N}\}$ are auxiliary variables. Upon deriving the solution,

for a query input $\boldsymbol{\phi}$, the center of the interval $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\phi}$ can be used as the *point estimate* of output, with (6) or (7) as its confidence interval. Notice that (8) and (9) are all convex optimization problems.

2.2 Robust IPM

In formulation IPM_N, the desired interval covers all observed data, including bad ones. However, in many practical situations, some observed samples may show patterns that are significantly different from the others, i.e. outliers. If we still choose all data for modeling, the interval tends to be over-conservative, which is an unwanted result. Because of the considerable influence of outliers on the identification performance, some robustness shall be imposed on the model by purposely weakening the effect of potential outliers. The most straightforward way to deal with outliers is to just discard them, which has already been extensively used in system identification, see e.g. Bai et al. (2002), Crespo et al. (2014) and Jaulin and Walter (2002). On the other hand, there are indeed some occasions where a high reliability is not so important, and the decision maker may prefer a tighter interval for a reduced reliability (e.g. 0.6 or 0.7 is enough) even if there are no outliers at all. In this case, discarding some samples are beneficial since the correlation between input and output can be informatively exhibited with a tight prediction interval.

Here in this paper, we carry out IPM identification allowing a fraction of data samples to be discarded, through which a narrower interval can be obtained to characterize the region where data are densely distributed and yield robustness against outliers. Suppose that n out of N samples are to be discarded, which shall be optimally selected by the model. This leads to the following robust counterpart of IPM_N:

$$\begin{aligned}\mathrm{IPM}_{N,n}: \quad \min_{F,\mathcal{I}_{N,n}} \quad & \mathbb{E}_{\boldsymbol{\phi}}\{\text{Size of } F(\boldsymbol{\phi})\} \\ \text{s.t.} \quad & y_i \in F(\boldsymbol{\phi}_i),\ \forall i \in \mathcal{I}_{N,n}, \\ & \mathcal{I}_{N,n} \triangleq \{i^{(1)},\dots,i^{(N-n)}\} \subseteq \mathbb{N}_{1:N},\end{aligned} \tag{10}$$

where $\mathcal{I}_{N,n}$ is a subset of $\mathbb{N}_{1:N}$ containing $N-n$ indices to be decided by the algorithm, whose complement involves samples that are discarded optimally.

However, due to its combinatorial nature, choosing $\mathcal{I}_{N,n}$ poses significant challenging in solving $\mathrm{IPM}_{N,n}$. One way is to resort to mixed-integer formulations including binary variables to indicate sample discarding (Luedtke et al. (2010)), where $\mathcal{I}_{N,n}$ can be optimized simultaneously with the set-valued mapping $F(\cdot)$. Alternatively, we can adopt the greedy algorithm based on solving the convex optimization problem IPM_N repeatedly. Every time discard one sample whose removal leads to the minimal interval width and repeat it for n times. Other methods, such as optimal removal proposed in Campi et al. (2009); Campi and Garatti (2011), can also be used.

3. A POSTERIORI RELIABILITY GUARANTEE

3.1 Reliability and a priori bound

By introducing the concept of discarding samples in IPM, it is expected that the output interval becomes narrower, and the prediction reliability decreases accordingly. A decision maker can be aware of the risk of using the derived interval for further uncertainty quantification if the reliability of an IPM is known, whose definition is given as follow:

Definition 2. Suppose the solution to IPM_N is F_N^*. The reliability $R(F_N^*)$ is defined as the probability that a new sample $(\boldsymbol{\phi}, y)$ is consistent with the derived set-valued mapping F_N^*:

$$R(F_N^*) = \mathbb{P}\{(\boldsymbol{\phi}, y) | y \in F_N^*(\boldsymbol{\phi})\}. \tag{11}$$

The above definition also applies to $\mathrm{IPM}_{N,n}$, and its reliability is denoted by $R(F_{N,n}^*)$. An assumption related to reliability is made below.

Assumption 3. (Existence and uniqueness) For every N and $0 \le n < N$ and every sample set $\{(\boldsymbol{\phi}_i, y_i) | i \in \mathbb{N}_{1:N}\}$, the solution of $\mathrm{IPM}_{N,n}$ exists and is unique.

Under the above assumption, we wish to derive a $(1-\varepsilon)$-reliability guarantee for $F_{N,n}^*$ such that it satisfies the following relationship:

$$\mathbb{P}^N\{R(F_{N,n}^*) \ge 1-\varepsilon\} \ge 1-\beta. \tag{12}$$

As aforesaid, a sample pair $(\boldsymbol{\phi}_i, y_i)$ can be seen as a random vector in some unknown probability space $\mathbb{P}$. Since the samples are all independently identically distributed, the whole sample set $\{(\boldsymbol{\phi}_i, y_i) | i \in \mathbb{N}_{1:N}\}$ is on the N-fold probability space $\mathbb{P}^N$. Under the Assumption 3, the optimal solution $F_{N,n}^*$ and the reliability $R(F_{N,n}^*)$ can be seen as relying on the sample set and thus both are also random set and random variable in $\mathbb{P}^N$. As a result, $R(F_{N,n}^*) \ge 1-\varepsilon$ is a random event associated with a probability, just as (12). If β decreases, the probability for $R(F_{N,n}^*) \ge 1-\varepsilon$ to happen will increase, and thus there is an interplay between ε and β. Achievements in such forms have been gained on the reliability of the IPM in literature. Reliability in both independent and weak-dependent situations are discussed in (Calafiore and Campi (2003)). An a priori bound for a special kind of IPM, minimax layers, is analyzed in Garatti et al. (2019). When the problem is convex, the discarding method and the a priori bound of the reliability after discarding is also mentioned in Campi and Garatti (2008); Campi et al. (2009); Campi and Garatti (2011), given as follow:

Theorem 4. (Campi et al. (2009)) Under Assumption 3, (12) holds with

$$\begin{aligned}\beta &= \beta_0 \sum_{i=0}^{n} \frac{(N-d)!}{(N-d-i)!i!} \cdot \frac{\varepsilon^i}{(1-\varepsilon)^i}, \\ \beta_0 &= \frac{N!}{(N-d)!d!}(1-\varepsilon)^{N-d},\end{aligned} \tag{13}$$

where d is the number of free parameters. In the above formulation, β can be determined with a given ϵ. For practical use, it is desirable attain the reliability guarantee $1-\varepsilon$ with a pre-specified confidence level. For instance, β is chosen as a very small positive number such as 10^{-6}. Based on the relationship between ε and β revealed in Theorem 4, the value of ε can be computed in a reverse order given the value of β, which essentially provides an estimate of upper ϵ-quantile of $F_{N,n}^*$ with $(1-\beta)\times 100\%$ confidence. An exclusive advantage of Theorem 4 its *distribution-free*

nature in that it holds for an arbitrary data-generating mechanism $\mathbb{P}$. Hence the case where $\mathbb{P}$ is unknown can be tackled desirably.

3.2 A posteriori bound

Note that the above a priori bound can be obtained before the sample set is obtained and the model is built, and thus is essentially solution-independent. It implies that equal values of ϵ will be attained for different realizations of $F^*_{N,n}$ provided that N and n are given, this bound remains the same, and thus cannot reveal the difference of reliability between different solutions. In fact, if the bound can be combined with some a posteriori information encoded in the solution and varies with sample sets, the deriving a posteriori bound after seeing $F^*_{N,n}$ could be more reliable.

Recent progress has been made on a posteriori bounds of scenario programs without sample discarding. The a posteriori bounds of reliability have been developed under both the convex (Campi and Garatti (2018)) and non-convex settings (Campi et al. (2018)). Shang and You (2020) proposed a posteriori bounds for reliability estimations combined with additional validation information. Here we extend a posteriori bounds to cases with sample discarding as a major contribution of this work. The construction relies on the concept of support sub-samples, whose definition is given below.

Definition 5. (Support sub-samples, Campi and Garatti (2018)) The support sub-samples of convex program IPM_N are defined as those whose removal does not affect the optimal solution.

Definition 6. (Support sub-samples with sample discarding) Suppose we have already got $\mathcal{I}^*_{N,n}$ after solving $\text{IPM}_{N,n}$. With $\mathcal{I}^*_{N,n}$ fixed, the optimization problem IPM_{N-n} reduces to a convex optimization problem, which is denoted by $\overline{\text{IPM}}_{N-n}$. Then the support sub-samples of $\text{IPM}_{N,n}$ are defined as those of $\overline{\text{IPM}}_{N-n}$.

It is worth mentioning that Definition 6 is new in literature, and Definition 5 given by Campi and Garatti (2018) does not applies directly to the case with sample discarding. Once discarded samples have been decided, $\text{IPM}_{N,n}$ reduces to $\overline{\text{IPM}}_{N-n}$, and their optimal solutions $F^*_{N,n}$ and $\overline{F}^*_{N-n}$ also coincide. Here we utilize useful information in the "decisive" support sub-samples of convex program $\overline{\text{IPM}}_{N-n}$ for reliability assessment of $\text{IPM}_{N,n}$, which is non-convex. It is known that for convex programs, the constraints of the support sub-samples are always active constraints, which become strict equalities at the optimal solution (Campi and Garatti (2018)). To establish an a posteriori guarantee, the following assumption is made.

Assumption 7. (Non-degeneracy, Campi and Garatti (2018)) With probability 1 the solution to IPM_N coincides with the solution defined by its support sub-samples only.

Denote by $s^*_{N,n}$ the number of support sub-samples of $\text{IPM}_{N,n}$. Now we are ready to put forward an a posteriori bound based on $s^*_{N,n}$ to give a guarantee on the reliability of prediction interval of $\text{IPM}_{N,n}$. Specifically, ε becomes a function of $s^*_{N,n}$, that is, $\varepsilon(s^*_{N,n})$, which can be determined after the problem is solved and $s^*_{N,n}$ is revealed. This explains why it is called an a posteriori bound. The main results are given as follows.

Theorem 8. Suppose that $\varepsilon_n(k),\ k = 0, 1, \ldots, u$ is a $[0, 1]$-valued function given by the user, $u = \min\{d, N-n\}$ and β is the optimal value of the following optimization problem:

$$\begin{aligned}\inf_{f(\cdot)\in C^u[0,1]}\ & f(1)\\ \text{s.t.}\quad & \frac{1}{k!}\frac{d^k}{dt^k}f(t) \geq \binom{N}{k}\binom{N-k}{n}t^{N-n-k}\\ & \qquad \cdot (1-t)^n \cdot \mathbf{1}_{[0,1-\varepsilon_n(k))}(t),\\ & \qquad\qquad \forall t \in [0,1],\ \forall\ k \in \mathbb{N}_{0:u},\\ & \frac{1}{k!}\frac{d^k}{dt^k}f(t) \geq 0,\ \forall t \in [0,1],\ \forall\ k \in \mathbb{N}_{u+1:d}.\end{aligned} \tag{14}$$

Under Assumption 7, it then follows that

$$\mathbb{P}^N\{R(F^*_{N,n}) \geq 1 - \varepsilon_n(s^*_{N,n})\} \geq 1 - \beta. \tag{15}$$

Now β is dependent on the selection of functionals $\varepsilon_n(\cdot)$. For convenience in practical use, the following theorem suggests a sound choice of constructing $\varepsilon_n(\cdot)$ reversely with β fixed.

Theorem 9. Define polynomial functions as:

$$\begin{aligned}&\varphi_{k,n}(t)\\ =&\frac{\beta}{N+1}\sum_{m=k}^{N-n}\binom{m}{k}t^{m-k} - \binom{N}{k}\binom{N-k}{n}t^{N-n-k}(1-t)^n,\\ &\qquad\qquad n \in \mathbb{N}_{0:N-1},\ k \in \mathbb{N}_{0:u}\end{aligned} \tag{16}$$

If $u = d$, define $t^0_{k,n}$ as the smallest root for $\varphi_{k,n}(t)$ in $[0, 1]$. It holds that $\{\varepsilon_n(k) = 1 - t^0_{k,n}\}$ defines a group of functions that satisfy $\mathbb{P}^N\{R(F^*_{N,n}) \geq 1-\varepsilon_n(s^*_{N,n})\} \geq 1-\beta$. If $u = N - n$, for $k < N - n$, $\varepsilon_n(k)$ is defined the same as above. For $k = N - n$, $\varepsilon_n(N - n) = 1$.
Also the constructed $\varepsilon_n(k)$ satisfies the monotonicity with k and n, that is $\varepsilon_n(k) > \varepsilon_n(k-1),\ k \in \mathbb{N}_{1:u},\ n \in \mathbb{N}_{0:N-1}$ and $\varepsilon_n(k) > \varepsilon_{n-1}(k),\ k \in \mathbb{N}_{0:u},\ n \in \mathbb{N}_{1:N-1}$.

Proof. The proofs of Theorems 8 and 9 can be made in a similar way as those in Campi and Garatti (2018) and are omitted due to space limitations.

To illustrate the characteristics of the established bounds, the following remarks are made.

Remark 10. Theorems 8 and 9 hold in a distribution-free manner, which can be used for reliability estimation of IPMs without knowing the true data-generating mechanism $\mathbb{P}$.

Remark 11. To use the proposed reliability guarantee, one has to first solve $\text{IPM}_{N,n}$ and then get $\mathcal{I}^*_{N,n}$ and $\overline{\text{IPM}}_{N-n}$. Afterwards, $s^*_{N,n}$ can be computed based on a greedy method by removing samples one by one from $\overline{\text{IPM}}_{N-n}$ and checking the solution (Campi et al. (2018)). In fact, the algorithmic efficiency can be improved by only searching among *active constraints*, since the constraints of the support sub-samples are always active in convex programs. Hence it suffices to restrict attentions to constraints with non-zero dual variables in $\overline{\text{IPM}}_{N-n}$ as candidate support sub-samples.

Remark 12. It is known that the number of support sub-samples of a convex program cannot exceed d (Calafiore

and Campi (2005)), provided that $d < N - n$. This is a standing assumption in literature, indicating that sufficient samples are used for problem formulation. However, it may be the case that $d \geq N - n$ where there are no adequate samples and the input dimensionality is high, which calls for the use of an extra *regularization term* to alleviate model over-fitting (Ljung et al. (2019)). Different from most existing results, the proposed a posteriori bound also applies to this situation when $d \geq N - n$, as indicated by Theorems 8 and 9.

4. SYSTEM IDENTIFICATION CASE STUDIES

In this section, we investigate the use of the proposed guarantee in reliability estimation based the SDP formulation (9). A system with one-dimensional input x and one-dimensional output y is adopted in this case study for clear illustrations. The original system for producing the sample set is $y = \frac{10}{\sqrt{2\pi}} e^{-x^2/2} - \frac{10}{\sqrt{2\pi}}$. 300 input samples are taken uniformly in $[-1, 1]$. All the output y is added with a Gaussian noise $\mathcal{N}(0, 0.03^2)$. Meanwhile, an additional heavy noise term $\mathcal{N}(0, 0.3^2)$ is added to the output y with probability 0.01 and this can be seen as the procedure for generating exceptional data points. Note that we present the structure of the system only for the readers to repeat the case. When conducting the system identification experiment, we regard the system as a black box and use nothing but only the sample set. A sample set with $N = 300$ is extracted from the sample space. We use polynomials with order 0-4 as basis functions for nonlinear model identification, such that input x is projected into $\boldsymbol{\phi} = [1, x, x^2, x^3, x^4]$. The number of discarded samples is chosen as $n = 3$. A greedy algorithm is used to decide which sample should be discarded, and in each iteration a convex scenario program is solved with a candidate sample removed. The final modeling result is visualized in Fig. 1. The red circles denote the support sub-samples after discarding and the green stars are the samples that already have been discarded.

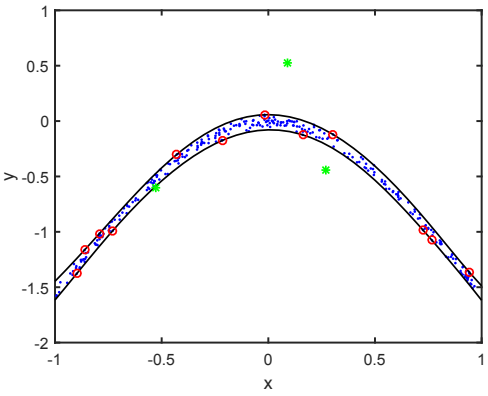

Fig. 1. Results of regression

In this case, we choose $\beta = 10^{-6}$. Recall that $N = 300, n = 3, d = 31$. After discarding $n = 3$ samples, the number of support sub-samples of the reserved 297 samples turns out to be $s^*_{300,3} = 12$. In virtue of Theorem 9, we can know $\varepsilon_3(12) = 0.1759$. So here comes the conclusion: $\mathbb{P}^{300}\{R(F^*_{300,3}) \geq 0.8241\} \geq 1 - 10^{-6}$. It means that with probability no less than $1 - 10^{-6}$, for an unknown input x, the probability for the real output y falling into the interval (7) is higher than 0.8241.

To testify the fidelity of the proposed a posteriori bound in lower-bounding the true reliability, we carry out extensive Monte Carlo simulations. 10^6 new samples are independently generated in the original sample space as a "practically true" mechanism for validation. For each IPM identified based on 300 random samples, the reliability of its interval is estimated as the percentage of samples falling into the interval in all 10^6 validation samples. For example, in a particular validation sample set derived, there are 94455 out of 10^6 samples that fall into the interval. So the empirical reliability is about 0.9446. As we can see, the lower bound we get by using theoretical research is 0.8241, which effectively bounds the empirical reliability. In comparison, the a priori bound in Theorem 4 is only 0.4491. Obviously the proposed a posteriori bound has much lighter conservatism than a priori bound.

Note that this is just the result based on one instance of sampling $N = 300$ samples. Next we carry out Monte Carlo experiments by repeating the above procedure 1,000 times and make a comparison between the empirical reliability and different probabilistic guarantees, as shown in Fig. 2.

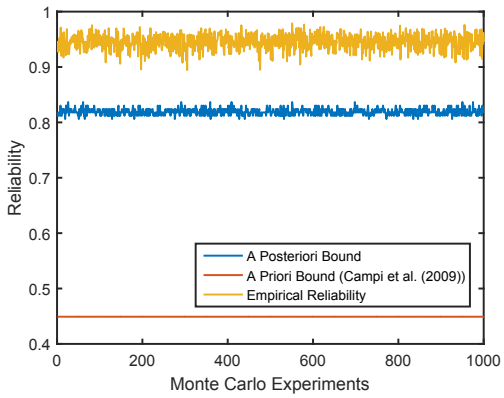

Fig. 2. The comparison of different guarantees of interval reliability under 1,000 Monte Carlo experiments

In 1,000 Monte Carlo experiments, samples are all independently sampled from the whole probability space. As a result, the empirical reliability is random. The a posteriori bound is also random because $s^*_{N,n}$ depends on the sample set. Since the a priori bound does not relies on the particular relaxation of datasets, in 1,000 Monte Carlo experiments the a priori guarantee in Campi et al. (2009) remains the same all the time. Although it successfully lower-bounds the true reliability, significant conservatism can be observed. In contrast, the proposed is always lower than the empirical reliability and is much higher than the a priori bound. The relationships of the a posteriori bound and the a priori bound with support sub-samples are further depicted in Fig. 3. With the increase of the number of support sub-samples s, the a posteriori bound decreases while the a priori bound remains the same. In this case, the a posteriori bound is always less conservative than the a priori bound for all possible realizations of s.

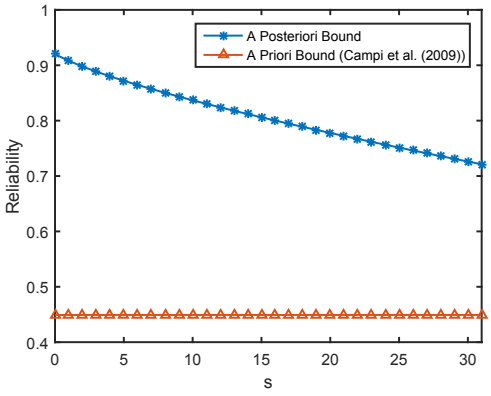

Fig. 3. Comparison between a priori and a posteriori guarantees with different numbers of support subsamples s

5. CONCLUSION

This paper establishes a probabilistic reliability guarantee for confidence interval obtained by IPM identification with sample discarding, which holds irrespective of the true data-generating mechanism. Compared to the a priori bound in literature, the proposed a posteriori bound is solution-dependent and turns out to be less conservative. Together with robust IPM and the a posteriori guarantee, one has access to more accurate information about model risk and reliability, which is particularly beneficial for further uncertainty quantification and decision-making.

REFERENCES

Bai, E.W., Cho, H., Tempo, R., and Ye, Y. (2002). Optimization with few violated constraints for linear bounded error parameter estimation. *IEEE Transactions on Automatic Control*, 47(7), 1067–1077.

Box, G.E., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Calafiore, G. and Campi, M.C. (2003). A learning theory approach to the construction of predictor models. *Discrete and Continuous Dynamical Systems*, supplement, 156–166.

Calafiore, G. and Campi, M.C. (2002). Interval predictors for unknown dynamical systems: An assessment of reliability. In *Proceedings of the IEEE Conference on Decision and Control*, volume 4, 4766–4771.

Calafiore, G. and Campi, M.C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1), 25–46.

Campi, M.C., Calafiore, G., and Garatti, S. (2009). Interval predictor models: Identification and reliability. *Automatica*, 45(2), 382–392.

Campi, M.C. and Garatti, S. (2008). The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3), 1211–1230.

Campi, M.C. and Garatti, S. (2011). A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2), 257–280.

Campi, M.C. and Garatti, S. (2018). Wait-and-judge scenario optimization. *Mathematical Programming*, 167(1), 155–189.

Campi, M.C., Garatti, S., and Ramponi, F.A. (2018). A general scenario theory for nonconvex optimization and decision making. *IEEE Transactions on Automatic Control*, 63(12), 4067–4078.

Chryssolouris, G., Lee, M., and Ramsey, A. (1996). Confidence interval prediction for neural network models. *IEEE Transactions on Neural Networks*, 7(1), 229–232.

Crespo, L.G., Giesy, D.P., and Kenny, S.P. (2014). Interval predictor models with a formal characterization of uncertainty and reliability. In *IEEE Conference on Decision and Control*.

Fravolini, Mario, L., Napolitano, Marcello, R., Core, Giuseppe, D., and Papa, U. (2018). Experimental interval models for the robust fault detection of aircraft air data sensors. *Control Engineering Practice*, 78, 196–212.

Garatti, S., Campi, M., and Carè, A. (2019). On a class of interval predictor models with universal reliability. *Automatica*, 110.

Heskes, T. (1997). Practical confidence and prediction intervals. In *Advances in Neural Information Processing Systems*, 176–182.

Jaulin, L., Kieffer, M., Braems, I., and Walter, E. (2001). Guaranteed non-linear estimation using constraint propagation on sets. *International Journal of Control*, 74(18), 1772–1782.

Jaulin, L. and Walter, E. (2002). Guaranteed robust nonlinear minimax estimation. *IEEE Transactions on Automatic Control*, 47(11), 1857–1864.

Ljung, L. (1999). *System Identification: Theory for The User*. PTR Prentice Hall, Upper Saddle River, NJ.

Ljung, L., Chen, T., and Mu, B. (2019). A shift in paradigm for system identification. *International Journal of Control*, 1–8.

Luedtke, J., Ahmed, S., and Nemhauser, G.L. (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2), 247–272.

Morita, H., Kase, T., Tamura, Y., and Iwamoto, S. (1996). Interval prediction of annual maximum demand using grey dynamic model. *International Journal of Electrical Power and Energy Systems*, 18(7), 409–413.

Shang, C., Huang, X., and You, F. (2017). Data-driven robust optimization based on kernel learning. *Computers & Chemical Engineering*, 106, 464–479.

Shang, C. and You, F. (2019a). A data-driven robust optimization approach to scenario-based stochastic model predictive control. *Journal of Process Control*, 75, 24–39.

Shang, C. and You, F. (2020). A posteriori probabilistic bounds of convex scenario programs with validation tests. *arXiv preprint arXiv:1903.11734*.

Shrestha, D.L. and Solomatine, D.P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, 19(2), 225–235.

Zhang, L., Luh, P.B., and Kasiviswanathan, K. (2003). Energy clearing price prediction and confidence interval estimation with cascaded neural networks. *IEEE Transactions on Power Systems*, 18(1), 99–105.