2021 60th IEEE Conference on Decision and Control (CDC)
December 13-15, 2021. Austin, Texas

Interval Predictor Models for Robust System Identification*

Luis G. Crespo†*, Sean P. Kenny†, Brendon K. Colbert†, and Tanner Slagel†

2021 60th IEEE Conference on Decision and Control (CDC) | 978-1-6654-3659-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CDC45484.2021.9683582

***Abstract*— This paper proposes a framework for the quantification of structured uncertainty in a plant model according to multivariable input-output data. The only restriction imposed upon such a model is for its outputs to depend continuously on the parameters. An Interval Predictor Model (IPM) prescribes the parameters of a computational model as a bounded, path-connected set thereby making each predicted output an interval-valued function of the inputs. The formulation proposed seeks the parameter set leading to the tightest enclosure of the data. This set, which is modeled as a semi-algebraic set having a tunable complexity level, enables the characterization of parameter dependencies commonly found in practice. This representation of the uncertain parameters makes the resulting plant model amenable to robustness analysis and robust control techniques based on polynomial optimization. Furthermore, scenario theory is used to evaluate the generalization properties of the identified model. This evaluation yields a formally-verifiable, distribution-free upper bound on the probability of future data falling outside the predicted output intervals.**

I. Introduction

Accurate uncertainty quantification in plant models is instrumental to the design of not overly conservative robust controllers. The classic approach to robust identification entails generating an uncertain plant model comprised of infinitely many deterministic plant models [1]. This is also the case for the Interval Predictor Models (IPM) [2], [3], which characterize the parameters of a computational model as a path-connected set thereby yielding predicted output intervals. The formulation proposed herein seeks the parameter set leading to the tightest enclosure of multivariable input-output data. Means to estimate IPMs for computational models having an affine parameter dependency are available [4]. This paper extends this framework to models having a continuous but otherwise arbitrary dependency on the parameters. This setting includes models having a nonlinear and implicit structure whose predictions require simulation or numerical integration.

The parameters of the plant model will be characterized as a set. A key attribute of this characterization is its compatibility with the computational model. For instance, if we are identifying the finite element model of a flexible structure, all the realizations of the uncertain plant model will correspond to a feasible pair of mass and stiffness matrices. This is in contrast to schemes for which the uncertainty model contains physically unrealizable plant models.

*This work was ostensibly supported by the Advanced Air Transport Technology Project from NASA.

†Dynamic Systems and Controls Branch, NASA Langley Research Center, Hampton, VA, 23681, USA.

*Corresponding author, luis.g.crespo@nasa.gov.

Parameters prescribed by a set are regarded as dependent when the range of any of them depends on the value taken by the others. Modeling these commonly found dependencies is instrumental in reducing unnecessary conservatism in the uncertain plant model, thereby reducing the required complexity of robust controllers designed for it [5]. The uncertain plant models resulting from the proposed framework are able to characterize such dependencies with varying levels of complexity.

II. Problem Statement

A dynamic system is postulated to act on an input variable, $x \in X \subset \mathbb{R}^{n_x}$, to produce an output variable, $y \in Y \subset \mathbb{R}^{n_y}$. Assume that n independent and identically distributed input-output pairs are drawn from an uncertain dynamical system, and denote by

$$\mathcal{D} \triangleq \{\delta^{(i)}\}_{i=1}^{n}, \tag{1}$$

where $\delta^{(i)} = (x^{(i)}, y^{(i)})$, the corresponding data sequence. Each element of this sequence will be called a datum or a scenario. Multiple scenarios are needed to account for the variability in the system response caused by measurement error, unmodeled/unmeasurable inputs, non-linearities, a changing environment, or by changes in the physical properties of an ensemble of nominally identical test articles. It is assumed hereafter that the effects of noise in the measured inputs and outputs have been filtered out using standard system identification techniques [6].

Furthermore, assume that a computational model of the underlying dynamical system is available. This input-output model will be given by the continuous function $y : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_y}$,

$$y(x, p), \tag{2}$$

where y is the predicted output given some input x and some parameter p.

The goals of this article are twofold. First, we want to identify an uncertain plant model according to $\mathcal{D}$. Instead of fitting all of the data by choosing a single parameter point, the thrust in this work is to characterize p as a set. This set is chosen such that the ensemble of predictions corresponding to all its elements tightly enclose the input-output data. This characterization makes each predicted output an interval-valued function of the input. Second, we want to assess the generalization properties of the identified model without actually modeling the underlying process that yielded $\mathcal{D}$.

A few remarks regarding notation are in order. A scenario might correspond to a single value or to a set of values. For instance, if each of the n input-output pairs corresponds to a different specimen in an ensemble of test articles, and

the data describes the Frequency Response Function (FRF) resulting from modal analysis, the input variable x will contain n_w frequencies, whereas the output variables y_1 and y_2 will contain the corresponding magnitude and phase. For simplicity in the notation we assume that all scenarios have the same input, so $x^{(i)} = \hat{x} \in \mathbb{R}^{n_w}$ for all $i = 1, \dots n$, and that each output has n_w elements, so $y^{(i)} \in \mathbb{R}^{n_w \times n_y}$. Specific components of the output will be referenced using subindices. For instance, the scalar $y^{(i)}_{j,k}$ will denote the j-th element of the k-th output of the i-th scenario. Furthermore, the data subsequence corresponding to the k-th output will be denoted as $\mathcal{D}_k$.

A classical model calibration problem consists in finding a point p in (2) leading to a predicted output that closely approximates the data. Techniques for identifying this parameter point are often cast as

$$p^\star(\mathcal{D}) = \underset{p \in G}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| y(\hat{x}, p) - y^{(i)} \right\|, \tag{3}$$

where $\| \cdot \|$ refers to a norm in the corresponding output space, and the feasible space, given by

$$G = \{p : g(p) \leq 0\}, \tag{4}$$

with $g : \mathbb{R}^{n_p} \to \mathbb{R}^{n_g}$, enforces additional performance and stability requirements.

III. Interval Predictor Models

The single-output $n_y = 1$ case will be considered first. The IPM corresponding to the input-output model (2), the input $x \in X$, and the set $\mathbb{P} \subset \mathbb{R}^{n_p}$ is defined as

$$I_y(y; x, \mathbb{P}) \triangleq \{y(x, p) \text{ for all } p \in \mathbb{P}\}. \tag{5}$$

Lemma 1: I_y is an interval when $y(x, p)$ is continuous in p for a fixed x, and $\mathbb{P}$ is closed, bounded, and path-connected. The proof is deferred to the Appendix. The conditions of Lemma (1) will be assumed to hold hereafter. The IPM in (5) can be written as

$$I_y(y; x, \mathbb{P}) = \left[\, \underline{y}(x, \mathbb{P}), \; \overline{y}(x, \mathbb{P}) \, \right], \tag{6}$$

where the functions

$$\underline{y}(x, \mathbb{P}) = \min_{p \in \mathbb{P}} y(x, p), \tag{7}$$

$$\overline{y}(x, \mathbb{P}) = \max_{p \in \mathbb{P}} y(x, p), \tag{8}$$

are the lower and upper IPM boundaries respectively. Therefore, an IPM is an interval-valued function of x given by the mapping of all possible values of p in $\mathbb{P}$ through the input-output model. For each $x \in X$, (5) is an interval in Y.

IPMs admit a functional interpretation: the graph of each member of the family of infinitely many output functions that results from evaluating (2) at all possible realizations of p in $\mathbb{P}$ and x in X lies between the lower and upper IPM boundaries and no tighter boundaries exist.

Equations (5) and (6) can be naturally extended to the multi-output/variable case $n_y > 1$. In this setting we have a collection of n_y interdependent IPMs given by $I_{y_k}(y_k; x, \mathbb{P})$, where $k = 1, \dots n_y$. This interdependency stems from all n_y IPMs having the same parameter set $\mathbb{P}$.

The formulation below seeks a set $\mathbb{P}$ of a given class leading to a data-enclosing IPM having a minimal width. This set will be characterized in the two-step process detailed below. The first step, described in Section IV, entails mapping each input-output datum onto a collection of parameter points. This mapping is built such that the ensemble of predictions corresponding to these points tightly enclose the datum. The second step, described in Section V, consists in finding a set that tightly encloses the points corresponding to all scenarios.

IV. Data Mapping

The multi-point IPMs corresponding to the parameter sequence $\mathcal{Q} = \left\{ q^{(j)} \right\}_{j=1}^{r}$, with $q^{(j)} \in \mathbb{R}^{n_p}$, are given by

$$\begin{aligned} I_{y_k}(y_k; x, \mathcal{Q}) &= \left[\min_{j=1\dots r} y_k \left(x, q^{(j)} \right), \; \max_{j=1\dots r} y_k \left(x, q^{(j)} \right) \right], \\ &= \left[\, \underline{y}_k(x, \mathcal{Q}), \; \overline{y}_k(x, \mathcal{Q}) \, \right], \end{aligned} \tag{9}$$

where $k = 1, \dots n_y$. The spread of the k-th single-output IPM in (9) is defined as

$$S_k(\mathcal{Q}) \triangleq \mathbb{E}_x \left[\overline{y}_k(x, \mathcal{Q}) - \underline{y}_k(x, \mathcal{Q}) \right], \tag{10}$$

where $\mathbb{E}_x[\cdot]$ is the expectation with respect to the input x in X, and k is fixed. The input distribution needed to compute (10), often prescribed by the experimentalist, renders $\hat{x}$. The collective spread of the IPMs in (9) is

$$S(\mathcal{Q}) = \sum_{k=1}^{n_y} \omega_k S_k(\mathcal{Q}), \tag{11}$$

where $\omega > 0$ is a fixed vector of weights.

Our goal is to identify the parameter sequence $\mathcal{Q}$ that minimizes $S(\mathcal{Q})$ while satisfying $\mathcal{D}_k \subseteq I_{y_k}(y_k; x, \mathcal{Q})$ for $k = 1, \dots n_y$. This sequence can be written as the union of n subsequences, each corresponding to a datum-containing IPM, i.e.,

$$\mathcal{Q} = \bigcup_{i=1}^{n} \mathcal{Q}^{(i)}, \tag{12}$$

where $\delta^{(i)} \subseteq I_y(y; x, \mathcal{Q}^{(i)})$. An algorithm for estimating $\mathcal{Q}^{(i)}$ from $\delta^{(i)}$ is presented next.

Algorithm 1 (Multi-point IPM): Set[1] $\mathcal{Q}^{(i)} \leftarrow p^\star(\delta^{(i)})$.

1) Evaluate

$$\underline{\lambda}^\star = \max_{\substack{j=1\dots n_w \\ k=1\dots n_y}} \; \underline{y}_k(\hat{x}_j, \mathcal{Q}^{(i)}) - y^{(i)}_{j,k},$$

and

$$\overline{\lambda}^\star = \max_{\substack{j=1\dots n_w \\ k=1\dots n_y}} \; y^{(i)}_{j,k} - \overline{y}_k(\hat{x}_j, \mathcal{Q}^{(i)}).$$

2) If $\underline{\lambda}^\star \leq 0$ and $\overline{\lambda}^\star \leq 0$ stop. Otherwise, go to Step 3.
3) Denote $j^\star$ and $k^\star$ as the j and k indices corresponding to the greatest value between $\underline{\lambda}^\star$ and $\overline{\lambda}^\star$.

[1] If several least squares solutions exist, compute an IPM for each of them, and chose the one with the smallest spread.

4) If $\underline{\lambda}^\star \geq \overline{\lambda}^\star$ solve

$$q^\star = \underset{p\in G}{\operatorname{argmin}} \left\{ S(A) : y_{k^\star}(\hat{x}_{j^\star}, p) \leq y^{(i)}_{j^\star,k^\star} \right\},$$

where $A = \mathcal{Q}^{(i)} \cup p$. Otherwise, solve

$$q^\star = \underset{p\in G}{\operatorname{argmin}} \left\{ S(A) : y_{k^\star}(\hat{x}_{j^\star}, p) \geq y^{(i)}_{j^\star,k^\star} \right\},$$

5) $\mathcal{Q}^{(i)} \leftarrow \mathcal{Q}^{(i)} \cup q^\star$. Go to Step 1.

Hence, $I_y(y; x, \mathcal{Q}^{(i)})$ is the multi-point IPM of minimal spread enclosing[2] $\delta^{(i)}$. The parameter points corresponding to all the elements in $\mathcal{D}$ comprise the sequence

$$\mathcal{P} \triangleq \left\{ \mathcal{Q}^{(i)} \right\}_{i=1}^{n}. \tag{13}$$

Therefore, the mapping $\delta^{(i)} \to \mathcal{Q}^{(i)}$ defined by Algorithm 1 yields the sequence of "surrogate data" in (13). A framework for estimating the set $\mathbb{P}$ in (5) from $\mathcal{P}$ is presented next.

V. Data-enclosing Sets

Any path-connected set enclosing the ensemble of parameter points $\mathcal{Q} = \{p^{(i)}\}_{i=1}^m$ comprising $\mathcal{P}$ yields a data-enclosing IPM, i.e., if $\mathcal{Q} \subset \mathbb{P}$ then $I_y(y; x, \mathcal{Q}) \subseteq I_y(y; x, \mathbb{P})$ for all $x \in X$. The strategies below seek semi-algebraic sets that maximize the tightness of such an enclosure.

To this end, consider the family of sets $\mathbb{P}(\theta)$, where θ is a parameterizing variable. The desired set is the member of this family corresponding to

$$\theta^\star(\mathcal{Q}) = \underset{\theta}{\operatorname{argmin}} \left\{ f(\mathbb{P}(\theta)) : \mathcal{Q} \subset \mathbb{P}(\theta) \right\}, \tag{14}$$

where $f(\mathbb{P}) \geq 0$ returns a value proportional to the size of $\mathbb{P}$. Hence, $\mathbb{P}(\theta^\star(\mathcal{Q}))$ is the element of the chosen class that encloses the surrogate data while having minimal size. The primal IPM $I_y(y; x, \mathcal{Q})$ and the resulting IPM $I_y(y; x, \mathbb{P}(\theta^\star(\mathcal{Q})))$ often differ, with the relationship between their spreads given by $S(\mathcal{Q}) \leq S(\mathbb{P}(\theta^\star(\mathcal{Q})))$.

As such, the desired set $\mathbb{P}$ will be designed to contain not only the elements of $\mathcal{Q}$ but also their neighborhood thereby making the identified model robust to uncertainty. The extent of this neighborhood should be set according to engineering judgement. Insufficiently large sets might yield to uncertain plant models that fail to accurately characterize the unknown system dynamics thereby incurring a possibly high risk, whereas the conservatism resulting from overly large sets might render unnecessarily complex controllers that perform poorly in practice.

Instantiations of the optimization program in (14) for $\mathbb{P}(\theta)$ families having various geometries, and different objective functions f are presented below. Each of the resulting sets $\mathbb{P}(\theta^\star(\mathcal{Q}))$ alone, as well as their intersection, complies with the path-connectedness property required by the IPM. This feature enables the analyst to adjust the level of robustness/conservatism in the identified plant model according to the particular application. When the additional requirements in (3) are present, $\mathbb{P}$ must be restricted further so it only contains elements of G in (4). This practice requires that G be path-connected as well.

A. Polytopes

The data-enclosing box of minimum volume is

$$\mathbb{P}(\theta^\star(\mathcal{Q})) = \left\{ p : \min_i p_j^{(i)} \leq p_j \leq \max_i p_j^{(i)},\ \forall j \right\}. \tag{15}$$

Even though this set is trivial to compute, it fails to model the parameter dependencies commonly present in $\mathcal{Q}$, thereby making most identified plant models overly conservative.

Another choice is the convex hull of $\mathcal{Q}$, given by

$$\mathbb{P}(\theta^\star(\mathcal{Q})) = \left\{ \sum_{i=1}^m \lambda_i\, p^{(i)} : \sum_{i=1}^m \lambda_i = 1,\ \lambda_i \geq 0,\ \forall i \right\}, \tag{16}$$

which is the data-enclosing convex polytope of minimum volume. This set can be computed using the Quickhull Algorithm [7].

B. Sum of Squares of Polynomials

Sliced-Normal (SN) distributions enable the characterization of multivariate data as a tight, data-enclosing semi-algebraic set. A summary of developments in [8] leading to this characterization is presented next. Consider the polynomial mapping from physical space p to feature space z given by $z = t(p; d)$, where $t : \mathbb{R}^{n_p} \times \mathbb{N} \to \mathbb{R}^{n_z}$ is the vector of monomials in variables p of positive degree less than or equal to d, so $n_z = \binom{n_p+d}{n_p} - 1$. A Sum of Squares (SOS) of polynomials in p of degree $2d$ is

$$\phi(z; \theta) = (z - \mu)^\top P (z - \mu), \tag{17}$$

where $\mu \in \mathbb{R}^{n_z}$, and $P \in S^{n_z}_{++}$ with S^n_{++} denoting the space of symmetric positive definite matrices in $\mathbb{R}^{n\times n}$. The joint density of a SN is

$$f_p(p; \theta) = \begin{cases} \frac{1}{c(\theta)} \exp\left(-\frac{1}{2}\phi(t(p; d); \theta)\right) & \text{if } p \in \Delta, \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

where $\theta = \{\mu, P\}$ is the shaping parameter, Δ is the support set, and $c(\theta)$ is the normalization constant. The super-level sets of (18) are

$$\mathbb{L}(\theta, d, \kappa) = \left\{ p \in \Delta : \phi(t(p; d), \theta) \leq \kappa \right\}, \tag{19}$$

where $\kappa \geq 0$. The maximization of the worst-case log-likelihood of the data, $\min_i \log f_p\left(p^{(i)}; \theta\right)$, yields

$$\theta^\star(\mathcal{Q}) = \underset{\theta}{\operatorname{argmax}} \min_{i=1,\dots m} l\left(p^{(i)}; \theta, d\right), \tag{20}$$

[2]The process of calculating $\mathcal{Q}$ by using Algorithm 1 is equivalent to solving the n optimization programs

$$\mathcal{Q}^{(i)} = \underset{\mathcal{R}\subset G}{\operatorname{argmin}} \left\{ S(\mathcal{R}) : \mathcal{D}_k^{(i)} \subset I_{y_k}(y_k; x, \mathcal{R}),\ k = 1, \dots n_y \right\},$$

for $i = 1, \dots n$, where $\mathcal{R}$ is a collection of parameter points. One could instead identify $\mathcal{Q}$ by solving the single optimization program

$$\mathcal{Q} = \underset{\mathcal{R}\subset G}{\operatorname{argmin}} \left\{ S(\mathcal{R}) : \mathcal{D}_k \subset I_{y_k}(y_k; x, \mathcal{R}),\ k = 1, \dots n_y \right\}.$$

This formulation, however, might lead to predictions that do not resemble individual scenarios, i.e., there might not be points p in $\mathcal{Q}$ for which $y^{(i)} \approx y(x^{(i)}, p)$, thereby making $\mathcal{Q}$ inadequate. A difficulty in solving these programs directly is not knowing the number of points in $\mathcal{R}$ needed for feasibility. In most cases this number is considerably smaller than $n_w n_y$ and $n_w n_y n$ respectively.

where l, the log-likelihood function, is

$$l(p;\theta,d) = -\frac{1}{2}\phi\big(t(p;d);\theta\big) - \log c(\theta). \tag{21}$$

The tightest data-enclosing element of (19) is given by

$$\mathbb{P}(\theta^\star(\mathcal{Q})) = \mathbb{L}\big(\theta^\star, d, \kappa\left(t(\mathcal{Q};d),\theta^\star\right)\big), \tag{22}$$

where

$$\kappa(\mathcal{R},\theta) \triangleq \max_i \Big\{\phi(r^{(i)};\theta) : r^{(i)} \in \mathcal{R}\Big\}. \tag{23}$$

It has been observed that (22) closely approximates the semi-algebraic set of degree $2d$ having minimal volume. The program in (20) is generally non-convex but for the two particular cases presented next it becomes convex.

1. Sliced-Normals of Polynomial Degree One

When $d = 1$ the set in (22) becomes an ellipsoid of volume $V(\mathcal{Q},\theta^\star)$, where

$$V(\mathcal{R},\theta) = \frac{\kappa(\mathcal{R},\theta)^{n_p}}{\det(P)}. \tag{24}$$

When $d = 1$ and $\Delta = \mathbb{R}^{n_p}$, the solution to (20) is given by the semi-definite program [8]

$$\min_{M\in S_{++}^{n_p+1},\,\alpha\in\mathbb{R}^+} \{\alpha - \log\det(P) : h(\mathcal{Q};M) \le \alpha\}, \tag{25}$$

where $h(p;M) = r^\top M r$ and $r = [1,p]^\top$. The relationship between M and $\theta = \{\mu, P\}$ is

$$M = \begin{bmatrix} \mu^\top P\mu & -\mu^\top P \\ -P\mu & P \end{bmatrix}. \tag{26}$$

Lemma 2: The set in (22) with $\theta^\star$ given by the solution to (25) is a data-enclosing ellipsoid of minimum volume.

The proof is deferred to the Appendix. Specialized algorithms for minimizing the volume of the data-enclosing ellipsoids in (41), such as Khachiyan's Algorithm [9], are more efficient than the SDP algorithms used to solve (25).

2. Sliced-Normals of Polynomial Degree Greater than One

SNs that maximize the worst-case likelihood of the data in feature space [8], as given by (20) for $c(\theta) = \sqrt{(2\pi)^{n_z}\det(P^{-1})}$, render a convex optimization program for any polynomial degree d. SNs of this class, whose elements are suboptimal relative to SNs calculated using the actual $c(\theta)$, should also be estimated by using the specialized algorithms mentioned above. Even though the resulting sets often enclose the data tighter than the minimum-volume ellipsoid, they might not be path-connected, thereby precluding their usage in (5).

Figure 1 shows data-enclosing sets of degrees three, four, five, and six for a dataset in $n_p = 2$ dimensions. The computational costs of calculating such sets using SDPT3 are 7.7s, 12.9s, 29.5s and 37.6s, whereas Khachiyan's Algorithm takes 4.5s, 3.8s, 3.0s and 2.7s respectively. The significant savings of Khachiyan's Algorithm increase rapidly with both n_p and d. Note that the volume of the sets do not decrease monotonically with d, and that all but the set corresponding to $d = 5$ are path-connected.

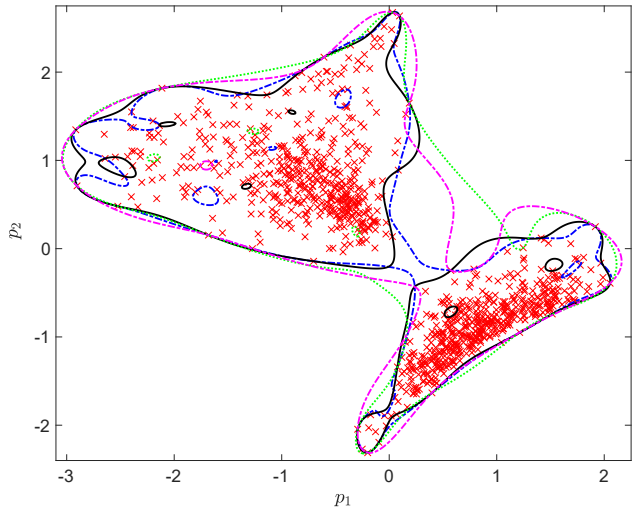

Fig. 1. Data-enclosing sets corresponding to $d = 3$ (magenta/dashed), $d = 4$ (green/dotted), $d = 5$ (black/solid), and $d = 6$ (blue/dashed-dotted). Elements of the dataset are shown (×).

The set (22) might not be path-connected thereby preventing its usage in (5). This property can be evaluated by using Bernstein polynomials within a branch and bound algorithm. However, the high computational cost of this practice renders it unsuitable when n_p is moderately large. This is also the case when the level sets of (18) are forced to be SOS-convex by adding constraints to (20). And even when this cost is acceptable, convexity is a more limiting requirement than path-connectedness thereby resulting in a suboptimal set, thus, in an unnecessarily wide IPM.

C. Multi-Ellipsoidal Sets

The formulation below seeks a set $\mathbb{P}$ given by the union of n_e intersecting ellipsoids thereby enforcing the desired path-connectedness property by design. The greater n_e the tighter the enclosure of the data, and therefore the smaller the corresponding IPM's spread.

To this end, consider the functions

$$s_j(p;\beta) = \begin{cases} 1 & \text{if } j = \underset{k}{\operatorname{argmin}}\ \phi\left(p;\{u^{(k)},N\}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

with $j = 1,\dots n_e$, where ϕ is in (17), and the parameter $\beta = \{\mathcal{C}, N\}$ is comprised of $\mathcal{C} = \{u^{(k)}\}_{k=1}^{n_e}$ with $u^{(k)} \in \mathbb{R}^{n_p}$, and $N \in S_{++}^{n_p}$. The functions in (27) act as a n_e-classifier for the m elements of $\mathcal{Q}$. In particular, (27) defines a Voronoi diagram in which the parameter space is partitioned into the n_e convex polytopes

$$\gamma_j(\beta) = \big\{p : s_j(p;\beta) = 1\big\}.$$

The members of this partition will be ordered such that γ_j shares a facet with γ_{j+1}. The corresponding partition of the data is $\{\mathcal{Q}_j\}_{j=1}^{n_e}$, where

$$\mathcal{Q}_j(\beta) = \{\, p : p \in (\mathcal{Q} \cap \gamma_j(\beta))\,\}.$$

This framework defines a n_e-means clustering algorithm in which β is the decision variable.

The multi-ellipsoidal set sought is given by

$$\langle \theta_j^\star, \beta^\star, \mathcal{V}^\star \rangle = \underset{\theta_j, \beta, \mathcal{V} \subset G}{\operatorname{argmin}} \sum_{j=1}^{n_e} V\Big(\mathcal{U}_j\big(\mathcal{Q}_j(\beta), \mathcal{V}\big), \theta_j\Big), \quad (28)$$

where V is defined in (24), $\theta_j = \{\mu_j, P_j\}$, $\mathcal{V} = \{v^{(k)}\}_{k=1}^{n_e-1}$ with $v^{(k)} \in \mathbb{R}^{n_p}$, and

$$\mathcal{U}_j(\mathcal{Q}_j, \mathcal{V}) = \mathcal{Q}_j \cup \Big\{ v^{(\max(j,\, n_e-1))}, v^{(\min(j,\, n_e-1))} \Big\}. \quad (29)$$

The roles played by various terms in the optimization program (28) are explained next. The objective function is the sum of the squares of the volume of n_e ellipsoids. At the optimum these ellipsoids are

$$\mathbb{L}_j = \mathbb{L}\Big(\theta_j^\star, 1, \kappa\left(\mathcal{U}_j(\mathcal{Q}_j(\beta^\star), \mathcal{V}^\star), \theta_j^\star\right)\Big), \quad (30)$$

where $j = 1, \ldots n_e$, and $\mathbb{L}$ is in (19). The ellipsoid $\mathbb{L}_j$ contains the elements of $\mathcal{Q}_j(\beta^\star)$ as well as the elements of $\mathcal{V}$ corresponding to the intersection between $\mathbb{L}_j$ and its neighboring ellipsoids, e.g., $v^{(1)}$ is in the intersection between $\mathbb{L}_1$ and $\mathbb{L}_2$, $v^{(2)}$ is in the intersection between $\mathbb{L}_2$ and $\mathbb{L}_3$, etc. These points are appended to $\mathcal{Q}$ in (29). The non-empty intersection between $\mathcal{U}_j$ and $\mathcal{U}_{j+1}$ for $j = 1, \ldots n_e - 1$ ensures that the union of the ellipsoids is path-connected. Lastly, $\mathcal{V} \subset G$ guarantees that at least one point at these intersections complies with the requirements in (3).

The parameter set $\mathbb{P}$ to be used by the IPM is

$$\mathbb{P}(\theta^\star(\mathcal{Q})) = \bigcup_{j=1}^{n_e} \mathbb{L}_j. \quad (31)$$

Figure 2 shows the sets in (31) corresponding to $n_e = 1$ and $n_e = 2$ for the same dataset in Figure 1. Note the sizable reduction in the volume of $\mathbb{P}$ caused by increasing n_e.

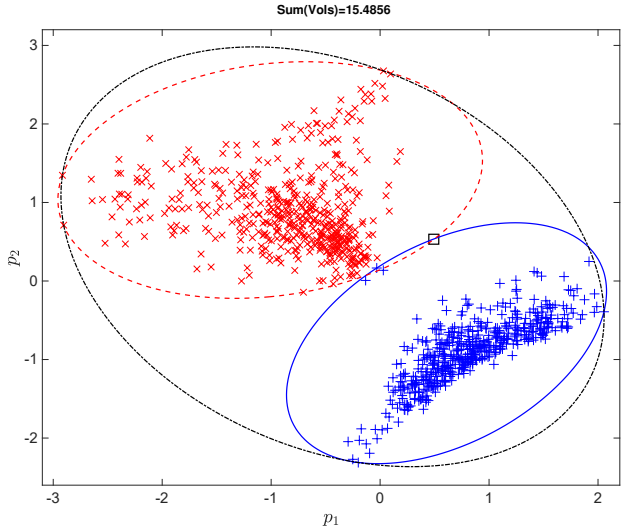

Fig. 2. Data-enclosing sets for $n_e = 1$ (black) and $n_e = 2$ (red and blue). The dataset $\mathcal{Q}$ is divided into $\mathcal{Q}_1(\beta^\star)$ ($\times$) and $\mathcal{Q}_2(\beta^\star)$ ($+$). The only element of $\mathcal{V}$ required in the latter case is shown as a square.

The optimization program in (28) is non-convex and NP-hard (as all k-means clustering algorithms are) thereby restricting its efficient application to moderately large n_p values. A suboptimal approximation to (28) corresponding to a fixed partition of the parameter space is found by solving the sequence of convex optimization programs detailed below. This partition, which is fully prescribed by β, can be found by using any clustering algorithm.

Algorithm 2 (Suboptimal multi-ellipsoidal set): Assume that the parameter space has been partitioned onto n_e subsets, that they are ordered such that consecutive elements are adjacent, and that $\mathcal{Q}_j$ contains the elements of $\mathcal{Q}$ falling onto the j-th subset.

1) Solve $\theta_j^\star(\mathcal{Q}_j) = \{\mu_j^\star, P_j^\star\}$ for $j = 1, \ldots n_e$.
2) Compute the symmetric matrix J, with

$$J_{ij} = V(p_{ij}, \theta_i^\star) + V(p_{ij}, \theta_j^\star), \quad (32)$$

for $i = 1, \ldots n_e$ and $j = 1, \ldots n_e$, where V is in (24), p_{ij} is the solution to

$$\min_{p \in G} \left\| p - \frac{q_{ij} V(\mathcal{Q}_i, \theta_i^\star) + q_{ji} V(\mathcal{Q}_j, \theta_j^\star)}{V(\mathcal{Q}_i, \theta_i^\star) + V(\mathcal{Q}_j, \theta_j^\star)} \right\|_2^2, \quad (33)$$

and

$$q_{kl} = \underset{p}{\operatorname{argmin}} \left\{ \phi(p; \theta_k^\star) : \phi(p; \theta_l^\star) \le \kappa(\mathcal{Q}_l, \theta_l^\star) \right\}.$$

3) Generate all the permutations of n_e elements having unique forward and backward orders[3]. Evaluate the cost of each of these permutations by summing the components of J for all overlapping pairs[4].
4) Find the permutation with the lowest cost. Expand the data partition by including the intersection points p_{ij} corresponding to all the pairs in this permutation[5]. Denote the resulting partition $\{\mathcal{Q}_j^\star\}_{j=1}^{n_e}$.
5) Repeat Step 1 using $\mathcal{Q}_j^\star$, and compute the corresponding data-enclosing ellipsoids $\mathbb{L}_j$ using (22).

Step 1 finds the ellipsoids that enclose each subset of the data. Step 2 evaluates the volume of pairs of ellipsoids each enclosing its corresponding data subsets while having a non-empty intersection. The solution to (33), whose calculation requires solving a pair of convex optimization programs upfront, is in this intersection. The program in (33) is convex when G is convex. Step 3 determines the sequence of intersecting ellipsoids having the smallest volume. Step 4 adds elements to the subsets of the data in order to guarantee that the resulting union of ellipsoids is path-connected. Finally, Step 5 calculates the desired ellipsoids.

The parameter set to be used by the IPM is given by (31), where the $\mathbb{L}_j$'s are given in Step 5.

VI. Examples

The robust system identification of a single-input single-output weakly nonlinear system having a non-collocated sensor actuator pair is considered next. The underlying physical

[3]e.g., all the permutations of interest for $n_e = 3$ are $s_1 = [1, 2, 3]$, $s_2 = [1, 3, 2]$, and $s_3 = [2, 1, 3]$. Note that $[2, 3, 1]$ is excluded because it coincides with the permutation that results from inverting the order of s_2.

[4]e.g., the cost of $[1, 2, 3]$ is $J_{12} + J_{23}$.

[5]e.g., the expanded partition for $[1, 2, 3]$ is $\mathcal{Q}_1^\star = \{\mathcal{Q}_1, p_{12}\}$, $\mathcal{Q}_2^\star = \{\mathcal{Q}_2, p_{12}, p_{23}\}$ and $\mathcal{Q}_3^\star = \{\mathcal{Q}_3, p_{23}\}$.

system consists of a spring-mass-damper configuration with 3 degrees of freedom. Each measured input corresponds to $n_w = 2500$ frequency points in the $X = [40, 500]$ rad/s range, and each measured output consists of the corresponding magnitude and phase. Averaging and windowing techniques are commonly used to mitigate the effects of measurement noise in such outputs. We seek to identify a robustly stable plant with transfer function

$$H(s,p) = \frac{p_1 s^4 + p_2 s^3 + p_3 s^2 + p_4 s + p_5}{s^6 + p_6 s^5 + p_7 s^4 + p_8 s^3 + p_9 s^2 + p_{10} s + p_{11}},$$

according to data sequences $\mathcal{D}$ of different lengths. Each scenario corresponds to an element of an ensemble of nominally identical test-articles.

Example 1: IPM for a Single Scenario

Figure 3 shows the data sequence $\mathcal{D} = \mathcal{D}^{(1)}$ used herein to identify the plant. First, Algorithm 1 was used to obtain $\mathcal{Q}^{(1)}$. The corresponding multi-point IPMs, shown in green, are comprised of 32 FRFs. A data-enclosing ellipsoidal set $\mathbb{P}(\theta^\star(\mathcal{Q}^{(1)}))$ was then calculated based on (41). The resulting IPM, $I_y(y; x, \mathbb{P}(\theta^\star(\mathcal{Q}^{(1)})))$, was simulated by using $n_s = 10000$ uniformly distributed samples in $\mathbb{P}(\theta^\star(\mathcal{Q}^{(1)}))$. The corresponding input-output functions lead to the multi-point IPM shown in blue. Both IPMs enclose the datum tightly while having practically equal spreads.

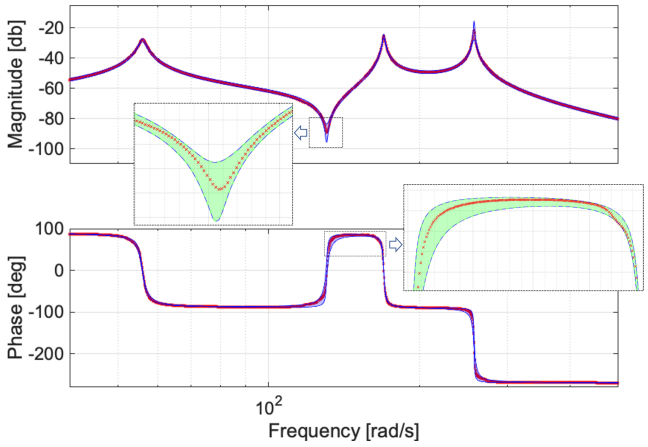

Fig. 3. The data sequence $\mathcal{D}^{(1)}$ (×), the multi-point IPM $I_y(y; x, \mathcal{Q}^{(1)})$ (green-shaded area), and the multi-point IPM corresponding to member functions of $I_y(y; x, \mathbb{P}(\theta^\star(\mathcal{Q}^{(1)})))$ (blue-shaded area).

Example 2: IPM for Multiple Scenarios

Next we identify a plant model according to a data sequence with $n = 100$ elements. Figure 4 shows the upper and lower envelopes of the data in red. The variability among the FRFs might be caused by variability in the material properties of an ensemble of test articles or by the effects of model-form uncertainty on a single test article, e.g., the power of the excitation in a modal experiment of a nonlinear structure changes the identified FRFs. The application of Algorithm 1 to each of the $n = 100$ data points yields a set $\mathcal{Q}$ with 4135 parameter points in $n_p = 10$ dimensions. The dispersion of these points, shown in red in Figure 5, indicate that they should be enclosed by a set that characterizes dependencies, thereby making (15) particularly unsuitable. The IPMs $I_y(y; x, \mathcal{Q})$, shown in Figure 4 as green-shaded areas, enclose the input-output data tightly.

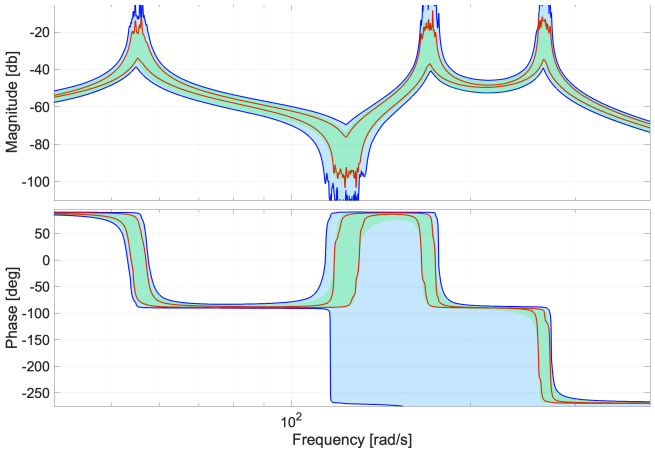

Fig. 4. Data envelopes (red), $I_y(y; x, \mathcal{Q})$ (green-shaded area), and multi-point IPM corresponding to 5000 member functions of $I_y(y; x, \mathbb{P}(\theta^\star(\mathcal{Q})))$ (blue-shaded area).

The set $\mathbb{P}$ chosen is the intersection of the ellipsoid in (22), the hyper-rectangle in (15), and the set of asymptotically stable plants. Figure 5 illustrates the tightness of the data enclosure by superimposing $n_s = 5000$ uniformly distributed points in $\mathbb{P}$. In contrast to this $\mathbb{P}$, the convex hull (16) could not be computed in finite time. The IPMs corresponding to these points are shown in Figure 4 as blue-shaded areas. The magnitude IPM encloses the data tightly, but the phase IPM does not. Figure 6 shows member FRFs of the phase IPM. Note that they split into two 360 deg apart branches at the first natural frequency. This feature, which is not caused by using an unwrapped angle, yield an IPM that might be considered overly wide and therefore unsuitable.

VII. Performance and Reliability Analyses

The smaller the spread of a data-enclosing IPM, the more informative the prediction, but the lower the probability of future data falling within its boundaries. These two conflicting notions, to be referred to as performance and reliability respectively, are studied next.

A. Performance Analysis

The developments that follow evaluate the degree by which subsets of $\mathbb{P}$ influence the spread of $I_y(y; x, \mathbb{P})$. The deviation of the member function $y_k(x, p)$ of $I_{y_k}(y; x, \mathbb{P})$ from the primal IPM $I_{y_k}(y; x, \mathcal{Q})$ is measured by

$$\lambda_k(p) = \frac{S_k(\mathcal{Q} \cup p) - S_k(\mathcal{Q})}{S_k(\mathcal{Q})}, \tag{34}$$

where $p \in \mathbb{P}$. The metric $\lambda_k(p)$ is the relative increase in the spread of the IPM that results from appending p to $\mathcal{Q}$. Hence, predictions for which $\lambda_k(p) = 0$ fall within $I_{y_k}(y; x, \mathcal{Q})$, whereas those for which $\lambda_k(p) > 0$ increase the

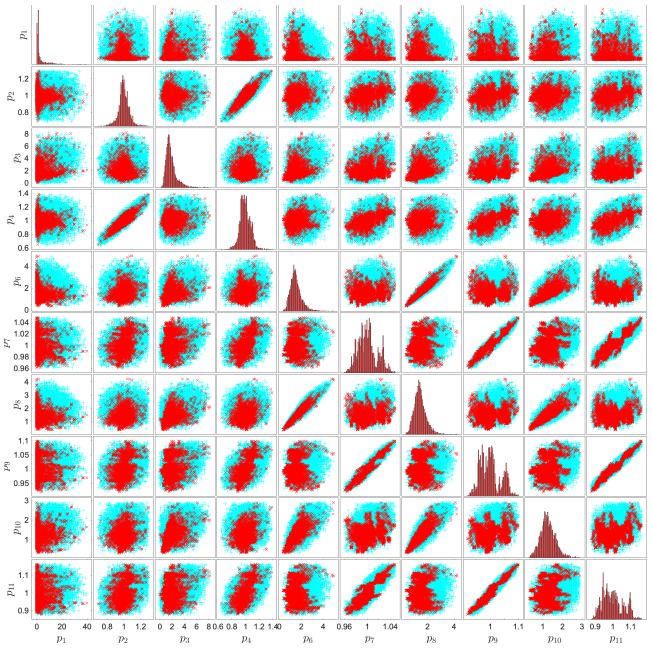

Fig. 5. Elements of $\mathcal{Q}$ ($\times$) and samples of $\mathbb{P}(\theta^\star(\mathcal{Q}))$ ($+$). Marginal densities are shown on the diagonal whereas off-diagonal subplots show 2-dimensional projections. p_5 has been omitted since it takes on a practically constant value. Values have been normalized to facilitate visualization.

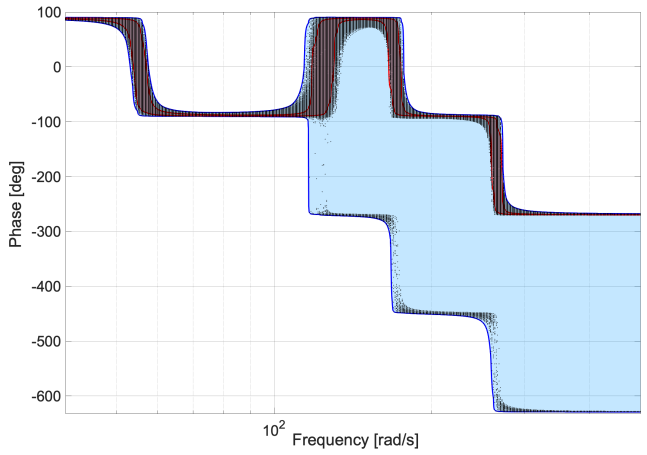

Fig. 6. Data envelopes (red lines), evaluations of member functions of $I_y(y;x,\mathbb{P}(\theta^\star(\mathcal{Q})))$ (black dots), and multi-point IPM (blue-shaded area).

spread proportionally to $\lambda_k(p)$. The cumulative distribution of $\lambda_k(p)$, F_{λ_k}, corresponding to a uniformly distributed p in $\mathbb{P}$ enables identifying parameter points responsible for an overly large spread. These points, which might occur rarely, i.e., $F^{-1}_{\lambda_k}(1-\epsilon) \ll F^{-1}_{\lambda_k}(1)$ for $0 < \epsilon \ll 1$, might be removed from $\mathbb{P}$ by imposing additional constraints to G in (4).

Example 3: Performance Analysis

The spread of the IPM in Example 2 is examined next. The evaluation of (34) at uniformly distributed samples in $\mathbb{P}$ led to values of λ_1 and λ_2 varying in the $[0, 0.035]$ and $[0, 10.5]$ ranges respectively. The upper limit of these intervals is consistent with the spreads seen in Figure 4. However, the empirical distribution F_{λ_2} shows that $F^{-1}_{\lambda_2}(0.98) = 0.05 \ll F^{-1}_{\lambda_2}(1) = 10.5$. Therefore, less than 2% of the volume of $\mathbb{P}$ causes the overly large phase spread. Tighter IPMs are obtained by not letting $\mathbb{P}$ enclose some of the worst-performing elements of $\mathcal{P}$, by further constraining $\mathbb{P}$, e.g., use $g(p) = \lambda_2(p) - 0.05$ in (4), or by increasing n_e. None of these practices, however, will be exemplified herein.

B. Reliability Analysis

This section presents means to formally evaluate the IPM's ability to characterize unseen data drawn from the same process by which $\mathcal{D}$ was obtained.

1. Background

Let $J : \Theta \to \mathbb{R}$ be a cost function of the decision variable $\theta \subset \mathbb{R}^{n_\theta}$, and h_δ be the set of decision points satisfying the system requirements for scenario $\delta \in \Delta$. Consider the scenario program

$$\theta^\star(\mathcal{P}) = \underset{\theta\in\Theta}{\operatorname{argmin}} \left\{ J(\theta) : \theta \in h_{\delta^{(i)}} \text{ for } i = 1, \dots n \right\}, \quad (35)$$

where $\mathcal{D} = \{\delta^{(i)}\}_{i=1}^n$ is drawn from a stationary but otherwise unknown process having a probability measure P. The violation probability of θ is defined as

$$V(\theta) \triangleq \mathsf{P}\left[\delta \in \Delta \mid \theta \notin h_\delta\right]. \quad (36)$$

The randomness of choosing a particular $\mathcal{D}$ out of the infinitely many having n scenarios makes $\theta^\star$, thus $V(\theta^\star)$, random. This notion can be quantified by using

$$\mathsf{P}^n\left[V(\theta^\star) \le \epsilon\right] \ge 1 - \beta. \quad (37)$$

This equation states that the probability of $\theta^\star$ violating a requirement for future data is no greater than ϵ with probability $1 - \beta$. The term ϵ is called the reliability parameter, whereas β is called the confidence. In regard to the latter parameter, note that $\theta^\star$ is a random element that depends on n observations of P. Therefore, the violation probability can be greater than ϵ for some random observations but not for others, and β refers to the probability $\mathsf{P}^n = \mathsf{P} \times \cdots \times \mathsf{P}$ of observing one of those bad set of n samples.

Scenario theory enables evaluating (37) without making any assumption on P. When the scenario program (35) is convex and its solution is unique, ϵ can be obtained from

$$\binom{k+n_\theta-1}{k} \sum_{i=0}^{k+n_\theta-1} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \le \beta, \quad (38)$$

where $k < n - n_\theta$ is the number of scenarios one might choose to remove from $\mathcal{D}$ before $\theta^\star$ was calculated. If (35) is also non-degenerate, and $k = 0$, a tighter bound can be obtained [10]. This bound is $\epsilon = 1 - \hat{t}$, where $\hat{t}$ is the zero of the polynomial in $t \in [0, 1]$,

$$\frac{\beta}{n+1} \sum_{i=\ell}^{n} \binom{i}{\ell} t^{i-\ell} - \binom{n}{\ell} t^{n-\ell} = 0, \quad (39)$$

and $0 \le \ell \le n_\theta$. The number of support constraints of the scenario program (35), ℓ, defined as the number of elements in a sub-sample $\mathcal{D}' \subset \mathcal{D}$ satisfying $\theta^\star(\mathcal{D}) = \theta^\star(\mathcal{D}')$, is a

complexity measure of $\theta^\star$. When the scenario program (35) is non-convex [11], ϵ is given by

$$\epsilon(\ell, n, \beta) = \begin{cases} 1 & \text{if } \ell = n \\ 1 - \left(\frac{\beta}{n\binom{n}{\ell}}\right)^{\frac{1}{n-\ell}} & \text{otherwise,} \end{cases} \quad (40)$$

where $0 \leq \ell \leq n$. Non-convex programs might admit several irreducible support sets, with the set having minimal cardinality yielding the tightest bound ϵ.

2. Reliability of IPMs

The smaller the probability of future data falling outside the IPM $I_y(y; x, \mathbb{P}(\theta^\star(\mathcal{Q})))$ the better its reliability. The means required to bound this probability depend on how $\theta^\star$ is computed. This process entails first applying Algorithm 1 to each of the n elements of $\mathcal{D}$ in order to obtain the $\mathcal{Q}^{(i)}$s comprising $\mathcal{P}$, and then finding a set $\mathbb{P}$ that optimally encloses the points in $\mathcal{P}$. The first step may be regarded as finding a non-convex map between each input-output datum, and a set of parameter points. Because this mapping is fixed, a scenario can be interpreted as either an element of $\mathcal{D}$ in (1), or as an element of $\mathcal{P}$ in (13). As such, the convexity of the program used to compute $\theta^\star$ from $\mathcal{P}$ determines whether convex or non-convex scenario theory is applicable, e.g., (38) and (39) apply to (41), whereas (40) applies to (31) and (22) for $d > 1$. Therefore, (37) becomes a probabilistic statement on $\mathcal{Q}^{(n+1)} \in \mathbb{P}$. On the other hand, $\mathcal{Q}^{(n+1)} \in \mathbb{P}$ implies $y^{(n+1)}(x^{(n+1)}) \in I_y(y; x, \mathbb{P})$. Hence, the data mapping $\delta^{(i)} \to \mathcal{Q}^{(i)}$ is immaterial from a reliability standpoint.

Determining the number of support constraints ℓ entails calculating $\theta^\star$ for clouds of parameter points corresponding to subsets of $\mathcal{P}$. It is noteworthy that this process does not require recalculating the $\mathcal{Q}^{(i)}$s. Further notice that the required value of ℓ is generally different from the value of ℓ corresponding to the scenario program having the enclosure of each parameter point as a constraint, e.g., the $2n_p$ points supporting a minimum-volume ellipsoid might correspond to the same input-output datum.

Example 4: Reliability Analysis

The reliability of the IPM in Example 2 is studied next. In this case ℓ can be calculated by solving n times the convex program (41), each time excluding an element of (13). In particular, the number of support scenarios is the number of times $\theta^\star(\mathcal{Q} \setminus \mathcal{Q}^{(i)})$ for $i = 1, \ldots n$ differ from $\theta^\star(\mathcal{Q})$. This practice yields $\ell = 21$, which along with $n = 100$, $\beta = 1\text{e}{-3}$ and (39) lead to $\epsilon = 0.4010$. Therefore, the probability of a new datum falling outside the IPM, $V(\theta^\star)$, is no greater than 0.4010 with confidence $1 - \beta = 0.999$.

3. Performance and Reliability Trade-off

The balance between high performance and high reliability can be formally traded-off using the above framework. For instance, if $\mathbb{P}$ is comprised of n_e ellipsoids, the evaluation of the performance function $S\left(\mathbb{P}(\theta^\star(\mathcal{Q}))\right)/S(\mathcal{Q})$, as given by (11), and the reliability function $1 - \epsilon(\theta^\star(\mathcal{P}))$ at various n_e values enables finding the desired balance.

APPENDIX

Proof of Lemma 1: $\mathbb{P}$ is path-connected if for every pair of points p_a and p_b in $\mathbb{P}$ there exists a continuous function $f : [0, 1] \to \mathbb{P}$ such that $f(0) = p_a$ and $f(1) = p_b$. Suppose $n_y = 1$, y is continuous in p, and $\mathbb{P}$ is path-connected. Let y_a, $y_b \in I_y$. There are $p_a, p_b \in \mathbb{P}$ such that $y_a = y(x, p_a)$ and $y_b = y(x, p_b)$. Since $\mathbb{P}$ is path-connected there is a continuous function $f : [0, 1] \to \mathbb{P}$ such that $f(0) = p_a$ and $f(1) = p_b$. Define $g : [0, 1] \to I_y$, where $g(t) = y(x, f(t))$ for each $t \in [0, 1]$. Since f and y are continuous, and the composition of two continuous functions is continuous, g is continuous. Since $g(0) = x_a$ and $g(1) = x_b$, I_y is path-connected. Being a path-connected subset of $\mathbb{R}$ is equivalent to being an interval. However this interval could be open, closed, half open, half closed, or unbounded. To ensure that I_y is bounded and closed, it is required that y is bounded and $\mathbb{P}$ is compact. □

Proof of Lemma 2: program (25) is equivalent to

$$\min_{\mu,\, P,\, \alpha \in \mathbb{R}^+} \left\{\alpha - \log\det(P) : \phi(\mathcal{Q}; \{\mu, P\}) \leq \alpha\right\}.$$

The change of variable $R = P/\alpha$ yields

$$\min_{\mu,\, R,\, \alpha} \{\alpha - n_p \log(\alpha) - \log\det(R) : \phi(\mathcal{Q}; \{\mu, R\}) \leq 1\},$$

whose solution is given by $\alpha^\star = n_p$, and

$$\min_{\mu,\, R \in S^{n_p}_{++}} \left\{ -\log\det(R) : \phi(\mathcal{Q}; \{\mu, R\}) \leq 1\right\}.$$

This optimization program is equivalent to

$$\min_{A \in S^{n_p}_{++},\, b \in \mathbb{R}^{n_p}} \left\{ \log\det(A^{-1}) : \|Ap^{(i)} - b\|_2 \leq 1,\, \forall i\right\}, \quad (41)$$

where $A = R^{1/2}$ and $b = R^{1/2}\mu$, which according to [12] yields a data-enclosing ellipsoids having minimal volume. □

REFERENCES

[1] D. K. de Vries and P. V. Hof, "Quantification of model uncertainty from data," *International journal of robust control*, vol. 4, no. 2, 1994.

[2] M. Campi, G. Calafiore, and S. Garatti, "Interval predictor models: Identification and reliability," *Automatica*, vol. 45, no. 2, 2009.

[3] S. Garatti, M. Campi, and A. Care, "On a class of interval predictor models with universal reliability," *Automatica*, vol. 110, Dec 2019.

[4] L. Crespo, S. Kenny, and D. Giesy, "Interval predictor models with a linear parameter dependency," *ASME Journal of verification, validation and uncertainty quantification*, vol. 1, no. 2, pp. 1–10, 2016.

[5] L. Crespo, D. Giesy, S. Kenny, and J. Deride, "A scenario optimization approach to system identification with reliability guarantees," *American Control Conference*, 2019.

[6] E. Morelli, "Estimating noise characteristics from flight test data using optimal Fourier smoothing," *Journal of Aircraft*, vol. 32, no. 4, 1995.

[7] C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, 1996.

[8] L. Crespo, B. Colbert, S. Kenny, and D. Giesy, "On the quantification of aleatory and epistemic uncertainty using sliced-normal distributions," *Systems & Control Letters*, vol. 134 (104560), 2019.

[9] M. J. Todd and E. A. Yildirim, "On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids," *Discrete Applied Mathematics*, vol. 155, no. 13, pp. 1731–1744, 2007.

[10] M. Campi and S. Garatti, "Wait-and-judge scenario optimization," *Mathematical Programming*, vol. 167, no. 1, 2018.

[11] ——, "A non-convex scenario theory," *IEEE Transactions in Automatic Control*, vol. 63, no. 12, 2018.

[12] N. Moshtagh, "Minimum volume enclosing ellipsoid," *Convex optimization*, vol. 111, no. 1 (January), pp. 1–9, 2005.