University of Groningen

Behavioral uncertainty quantification for data-driven control

Padoan, Alberto; Coulson, Jeremy; Van Waarde, Henk J.; Lygeros, John; Dorfler, Florian

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Behavioral uncertainty quantification for data-driven control

Alberto Padoan, Jeremy Coulson, Henk J. van Waarde, John Lygeros, and Florian Dörfler

Abstract— This paper explores the problem of uncertainty quantification in the behavioral setting for data-driven control. Building on classical ideas from robust control, the problem is regarded as that of selecting a metric which is best suited to a data-based description of uncertainties. Leveraging on Willems' fundamental lemma, restricted behaviors are viewed as subspaces of fixed dimension, which may be represented by data matrices. Consequently, metrics between restricted behaviors are defined as distances between points on the Grassmannian, i.e., the set of all subspaces of equal dimension in a given vector space. A new metric is defined on the set of restricted behaviors as a direct finite-time counterpart of the classical gap metric. The metric is shown to capture parametric uncertainty for the class of autoregressive (AR) models. Numerical simulations illustrate the value of the new metric with a data-driven mode recognition and control case study.

I. INTRODUCTION

In a typical control design problem, the role of data (time series) has been long dictated by *indirect approaches* [1, 2], where system identification is sequentially followed by model-based control. However, the advent of large data sets and the ever-increasing computing power, combined with the ongoing revolution brought about by machine learning methodologies, has recently triggered a renewed appreciation for *direct approaches*, where the objective is to infer optimal decisions directly from measured data.

A cornerstone of this newly emerging trend in control is a far-reaching result due to Willems and co-authors [3], commonly known as the *fundamental lemma*. Leveraging on the behavioral approach to system theory [4, 5], the fundamental lemma establishes that parametric models of a data-generating linear time-invariant (LTI) system may be replaced by a raw data matrix time series, provided the dynamics are sufficiently excited. Following the contributions [6–8], the number of new data-driven control algorithms has boomed over the past few years, see, e.g., [9] for a recent overview. A convincing demonstration of the potential of direct approaches to data-driven control is the successful implementation of the DeePC algorithm [6] in a wide range of experimental case studies, including synchronous motor drives [10], grid-connected power converters [11], and aerial robotics [12].

The new wave of data-driven control algorithms has primarily modeled uncertainty by ellipsoids [12–17]. While effective in many circumstances, this approach disregards the geometric structure of the data, leading to a possibly coarse characterization of uncertainty.

This paper explores the problem of uncertainty quantification in data-driven control of LTI systems. We seek a data-based behavioral description of uncertainty. Building on the rich legacy of robust control theory [18–21], we identify the problem of uncertainty quantification with that of selecting a "natural" metric to study robustness questions. The starting point of our analysis is a seemingly elementary, yet profound consequence of the fundamental lemma: restricted behaviors may be regarded as subspaces of fixed dimension and represented directly by data matrices. Building on this premise, we identify restricted behaviors with points on the *Grassmannian* $\mathrm{Gr}(k, N)$, *i.e.*, the set of all subspaces of dimension k in $\mathbb{R}^N$, endowed with the structure of a (quotient) manifold. The L-gap metric is then introduced as a direct finite-time counterpart of the classical gap metric [18–24], which measures the distance between graphs of input-output operators and allows one to compare the closed-loop behavior of different systems subject to the same feedback controller.

Contributions: The contributions of the paper are fourfold: (i) we define a new (representation free) metric on the set of restricted behaviors; we show that this metric is easily computed via measured data and readily understood as a distance between trajectories; (ii) we show that our metric can be used for uncertainty quantification for behaviors described by AR models; (iii) we connect the L-gap to the classical gap metric on ℓ_2 from robust control theory; and (iv) we demonstrate the benefits brought by the L-gap in a data-driven mode recognition and control case study.

Paper organization: The remainder of this paper is organized as follows. Section II provides basic definitions regarding behavioral systems. Section III introduces a new metric between restricted behaviors, which is then used for uncertainty quantification purposes and shown to be closely connected to the classical gap metric on ℓ_2. Section IV illustrates the theory with a numerical case study. Section V provides a summary of the main results and an outlook to future research directions. The proofs of our main results can be found in [25].

Notation: The set of positive and non-negative integers are denoted by $\mathbb{N}$ and $\mathbb{Z}_+$, respectively. The set of positive integers $\{1, 2, \ldots, p\}$ is denoted by $\mathbf{p}$ for every $p \in \mathbb{N}$. The set of real numbers is denoted by $\mathbb{R}$. The transpose, image, and kernel of the matrix $M \in \mathbb{R}^{p \times m}$ are denoted by M^{T}, $\mathrm{im}\, M$, and $\ker M$, respectively. A map f from X to Y is denoted by $f : X \to Y$; $(Y)^X$ denotes the collection of all such maps. The *backward t-shift* denoted by σ^t is defined as $(\sigma^t f)(t') = f(t + t')$ for all $t, t' \in \mathbb{Z}_+$.

A. Padoan, J. Coulson, J. Lygeros, F. Dörfler are with the Department of Information Technology and Electrical Engineering at ETH Zürich, Zürich, Switzerland {apadoan, jcoulson, lygeros, dorfler}@control.ee.ethz.ch. H. J. van Waarde is with the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands. h.j.van.waarde@rug.nl.

II. Behavioral Systems

A. Preliminaries in behavioral system theory

Following [5], we introduce some basic notions and results on behavioral systems.

Definition 1. A *dynamical system* Σ is a triple $\Sigma = (\mathbb{Z}_+, \mathbb{R}^q, \mathcal{B})$, where $\mathbb{Z}_+$ is the *time set*, $\mathbb{R}^q$ is the *signal space*, and $\mathcal{B} \subseteq (\mathbb{R}^q)^{\mathbb{Z}_+}$ is the *behavior* of the system.

Definition 2. A dynamical system $\Sigma = (\mathbb{Z}_+, \mathbb{R}^q, \mathcal{B})$ is *linear* if $\mathcal{B}$ is a linear subspace of $(\mathbb{R}^q)^{\mathbb{Z}_+}$, *time invariant* if $\sigma^t(\mathcal{B}) \subseteq \mathcal{B}$ for all $t \in \mathbb{Z}_+$, and *complete* if $\mathcal{B}$ is closed in the topology of pointwise convergence.

The structure of an LTI dynamical system is characterized by a set of integer invariants known as *structure indices* [4, Section 7]. The most important ones are the *number of inputs (or free variables)* m, the *lag* l, and the *order* n. The structure indices are intrinsic properties of a dynamical system, as they do not depend on its representation. The *complexity* of a dynamical system is defined as $c = (m, l, n)$. The class of all complete linear, time invariant systems (with complexity c) is denoted by $\mathcal{L}^q$ ($\mathcal{L}^{q,c}$). By a convenient abuse of notation, we shall also write $\mathcal{B} \in \mathcal{L}^q$ ($\mathcal{B} \in \mathcal{L}^{q,c}$).

Definition 3. Let $\mathcal{B} \in \mathcal{L}^q$ and $T \in \mathbb{N}$. The *restricted behavior (in the interval $[1, T]$)* is the set $\mathcal{B}|_T = \{w = (w_1, \ldots, w_T) \in \mathbb{R}^{qT} \mid \exists v \in \mathcal{B} : w_t = v_t, \forall t \in \mathbf{T}\}$. A vector $w \in \mathcal{B}|_T$ is a *T-length trajectory of the dynamical system $\mathcal{B}$*.

The following lemma characterizes the dimension of a restricted behavior $\mathcal{B}|_L \in \mathcal{L}^{q,c}$ in terms of its complexity.

Lemma 1. [26, Lemma 2.1] Let $\mathcal{B} \in \mathcal{L}^{q,c}$. Then $\mathcal{B}|_L$ is a subspace of $\mathbb{R}^{qL}$, the dimension of which is $\dim \mathcal{B}|_L = mL + n$, for $L > l$.

Definition 4. A dynamical system $\mathcal{B} \in \mathcal{L}^q$ is *controllable* if for every $T \in \mathbb{N}$, $w^1 \in \mathcal{B}|_T$, and $w^2 \in \mathcal{B}$ there exists $T' \in \mathbb{Z}_+$, and $w \in \mathcal{B}$ such that $w_t = w_t^1$ for $t \in \mathbf{T}$ and $w_t = w_{t-T-T'}^2$ for $t > T + T'$.

In other words, a dynamical system is controllable if any two trajectories can be patched together in finite time.

B. The fundamental lemma

Given a T-length trajectory $w \in \mathbb{R}^{qT}$ of a controllable dynamical system $\mathcal{B} \in \mathcal{L}^q$, one may obtain a non-parametric representation of the restricted behavior using a result first presented in [3], which over time became known as the *fundamental lemma*. To state this result, we introduce some preliminary notions.

Definition 5. The *Hankel matrix* of depth $L \in \mathbf{T}$ associated with $w \in \mathbb{R}^{qT}$ is defined as

$$H_L(w) = \begin{bmatrix} w_1 & w_2 & \cdots & w_{T-L+1} \\ w_2 & w_3 & \cdots & w_{T-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ w_L & w_{L+1} & \cdots & w_T \end{bmatrix} \in \mathbb{R}^{(qL) \times (T-L+1)}.$$

Definition 6. A vector $u \in \mathbb{R}^{mT}$ is *persistently exciting of order L* if $H_L(u)$ is full row rank, i.e., $\operatorname{rank} H_L(u) = mL$.

Persistency of excitation plays a key role in system identification and adaptive control [1, 2, 27]. A necessary condition for $u \in \mathbb{R}^{mT}$ to be persistently exciting of order L is that $H_L(u)$ has at least as many columns as rows, i.e., $T \geq T_{\min} = (m+1)L - 1$. We are now ready to state the fundamental lemma [3].

Lemma 2 (Fundamental lemma)**.** Consider a controllable dynamical system $\mathcal{B} \in \mathcal{L}^{q,c}$, with input/output partition $w = (u, y)$. Assume $w^d = (u^d, y^d) \in \mathcal{B}|_T$ and u^d is persistently exciting of order $L + n$. Then $\mathcal{B}|_L = \operatorname{im} H_L(w^d)$.

Lemma 2 is of paramount importance in data-driven control [28]. It provides conditions for the restricted behavior $\mathcal{B}|_L$ to be completely characterized by the image of the Hankel matrix $H_L(w^d)$. As a result, the subspace $\operatorname{im} H_L(w^d)$ can be regarded as a non-parametric representation of the dynamical system $\mathcal{B}$, so long as L-length trajectories are considered. The controllability and persistency of excitation assumptions can be removed by focusing on behaviors of fixed complexity and instead using the rank condition descending from Lemma 2

$$\operatorname{rank} H_L(w^d) = mL + n. \tag{1}$$

Lemma 3. [28, Corollary 19] Consider a dynamical system $\mathcal{B} \in \mathcal{L}^{q,c}$ and an associated T-length trajectory $w^d \in \mathcal{B}|_T$. For $L > l$, $\mathcal{B}|_L = \operatorname{im} H_L(w^d)$ if and only if (1) holds.

For convenience, in the sequel a T-length trajectory w^d of $\mathcal{B} \in \mathcal{L}^{q,c}$ is said to be *sufficiently excited of order L* if it satisfies the rank condition (1). All of these "low rank" results hold obviously for the *deterministic* LTI case, but they can also be used to design effective de-noising schemes by low-rank approximation or relaxations thereof [9].

III. A metric on restricted behaviors

This section explores the issue of uncertainty quantification using a data-based behavioral description of uncertainties. The starting point of our analysis is a seemingly elementary, yet profound consequence of the fundamental lemma: restricted behaviors may be regarded as subspaces of equal dimension, which may be represented directly by data matrices. Thus, restricted behaviors may be identified with points on the *Grassmannian* $\operatorname{Gr}(k, N)$, *i.e.*, the set of all subspaces of dimension k in $\mathbb{R}^N$, endowed with the structure of a (quotient) manifold [29, p.63]. Metrics between restricted behaviors thus arise from the underlying Grassmannian structure.

Proposition 1. The function d is a metric on the set of all restricted behaviors $\mathcal{B}|_L \in \mathcal{L}^{q,c}$, with $L > l$, whenever d is a metric on $\operatorname{Gr}(mL + n, qL)$.

Proof. Let $L > l$ and let d be a metric on $\operatorname{Gr}(mL + n, qL)$. By Lemma 1, the set of all restricted behaviors $\mathcal{B}|_L \in \mathcal{L}^{q,c}$ is a subset of $\operatorname{Gr}(mL + n, qL)$. Then the set of all restricted behaviors $\mathcal{B}|_L \in \mathcal{L}^{q,c}$ endowed with the metric d is also a metric space, since any subset of a metric space is itself a metric space with respect to the induced metric [30, p.38]. $\square$

With these premises, a natural question is: what is a good notion of distance for restricted behaviors? Ideally, a metric should be intrinsic, easily computed, and readily understood in system-theoretic terms. The aforementioned properties provide an identikit of the desired distance and pave the way for the discussion in this section, where we explore a notion of distance between restricted behaviors.

A. The gap between restricted behaviors

The gap metric plays a pivotal role in control theory [18 – 24] and, in many ways, it reflects the intuitive notion of distance between subspaces. This section introduces the L-gap metric as a direct finite-time counterpart of the classical gap metric. To this end, we recall a few preliminary notions.

Let $\mathcal{S}$ be a normed space with norm $\|\cdot\|$. Let $v \in \mathcal{S}$ and let $\mathcal{W}$ be subspace of $\mathcal{S}$. The *distance between v and $\mathcal{W}$* is defined as $\delta(v, \mathcal{W}) = \inf_{w \in \mathcal{W}} \|v - w\|$ [31, p.7]. If $\|\cdot\|_2$ is the Euclidean 2-norm in $\mathbb{R}^N$, the distance between v and $\mathcal{W}$ is the distance between v and its projection onto $\mathcal{W}$, *i.e.*, $\delta(v, \mathcal{W}) = \|(I - P_{\mathcal{W}})v\|_2$, where $P_{\mathcal{W}}$ is the orthogonal projector onto the subspace $\mathcal{W}$. Hence, the distance between a vector v of unit norm projected onto $\mathcal{V}$ and $\mathcal{W}$ is $\|P_{\mathcal{V}} - P_{\mathcal{W}}\|$. This motivates the following definition.

Definition 7. [19, p.30] Let $\mathcal{V}$ and $\mathcal{W}$ be closed subspaces of a Hilbert space $\mathcal{H}$. The *gap between $\mathcal{V}$ and $\mathcal{W}$* is defined as

$$\text{gap}_{\mathcal{H}}(\mathcal{V}, \mathcal{W}) = \|P_{\mathcal{V}} - P_{\mathcal{W}}\|, \tag{2}$$

where $P_{\mathcal{V}}$ and $P_{\mathcal{W}}$ are the orthogonal projectors onto $\mathcal{V}$ and $\mathcal{W}$, respectively.

The gap between $\mathcal{V}$ and $\mathcal{W}$ may be expressed as [19, p.30]

$$\text{gap}_{\mathcal{H}}(\mathcal{V}, \mathcal{W}) = \max \left\{ \|(I - P_{\mathcal{W}})P_{\mathcal{V}}\|, \|(I - P_{\mathcal{V}})P_{\mathcal{W}}\| \right\}.$$

In particular, $0 \leq \text{gap}_{\mathcal{H}}(\mathcal{V}, \mathcal{W}) \leq 1$ for all $\mathcal{V}$ and $\mathcal{W}$. To streamline the exposition, we also recall the notion of *directed gap between $\mathcal{V}$ and $\mathcal{W}$* which is defined as

$$\overrightarrow{\text{gap}}_{\mathcal{H}}(\mathcal{V}, \mathcal{W}) = \sup_{\substack{v \in \mathcal{V} \\ \|v\|=1}} \delta(v, \mathcal{W}) = \|(I - P_{\mathcal{W}})P_{\mathcal{V}}\|. \tag{3}$$

Clearly, $\text{gap}_{\mathcal{H}}(\mathcal{V}, \mathcal{W}) = \max\{ \overrightarrow{\text{gap}}_{\mathcal{H}}(\mathcal{V}, \mathcal{W}), \overrightarrow{\text{gap}}_{\mathcal{H}}(\mathcal{W}, \mathcal{V}) \}$. Note that no explicit mention to any particular choice of $\|\cdot\|$ is actually needed when the ambient Hilbert space $\mathcal{H}$ is $\mathbb{R}^N$, since all the gap functions are equivalent [32, p.91]. Throughout the paper, we consider the gap metric corresponding the Euclidean 2-norm $\|\cdot\|_2$ for simplicity.

We are now ready to introduce a notion of distance between restricted behaviors.

Definition 8. Let $\mathcal{B} \in \mathcal{L}^{q,c}$ and $\tilde{\mathcal{B}} \in \mathcal{L}^{q,c}$. For $L \in \mathbb{Z}_+$, the *L-gap between $\mathcal{B}$ and $\tilde{\mathcal{B}}$* is defined as

$$\text{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) = \text{gap}(\mathcal{B}|_L, \tilde{\mathcal{B}}|_L). \tag{4}$$

The *directed L-gap between $\mathcal{B}$ and $\tilde{\mathcal{B}}$* is defined as $\overrightarrow{\text{gap}}_L(\mathcal{B}, \tilde{\mathcal{B}}) = \overrightarrow{\text{gap}}(\mathcal{B}|_L, \tilde{\mathcal{B}}|_L)$.

The L-gap can also be defined for behaviors with different lags. However, for clarity of exposition we define it here for behaviors of the same complexity c.

1) System-theoretic interpretation. Given $\mathcal{B} \in \mathcal{L}^{q,c}$, consider the problem of estimating the closest trajectory $w \in \mathcal{B}|_L$ to a given measured trajectory $\tilde{w} \in \mathbb{R}^{qL}$ which belongs to a possibly distinct behavior $\tilde{\mathcal{B}} \in \mathcal{L}^{q,c}$, *i.e.*,

$$\begin{aligned} \underset{w \in \mathcal{B}|_L}{\text{minimize}} \quad & \|w - \tilde{w}\|_2^2, \\ \text{subject to} \quad & \tilde{w} \in \tilde{\mathcal{B}}|_L. \end{aligned}$$

By Lemma 1, $\mathcal{B}|_L$ is a subspace and the estimation error is

$$\inf_{w \in \mathcal{B}|_L} \|w - \tilde{w}\|_2 = \left\| (I - P_{\mathcal{B}|_L})\tilde{w} \right\|_2. \tag{5}$$

Now suppose $\tilde{\mathcal{B}}$ is known to be such that $\text{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) \leq \epsilon$. Then

$$\sup_{\substack{\tilde{w} \in \tilde{\mathcal{B}}|_L \\ \|\tilde{w}\|_2 \neq 0}} \inf_{w \in \mathcal{B}|_L} \frac{\|w - \tilde{w}\|_2}{\|\tilde{w}\|_2} \leq \epsilon.$$

In other words, $\text{gap}_L(\mathcal{B}, \tilde{\mathcal{B}})$ is an upper bound for the *worst case* relative estimation error. The domain of the L-gap metric may be extended to measure distances between subspaces of different dimension [33], so these results may be used, e.g., for tracking a reference or smoothing of a noisy trajectory $\tilde{w}$. We elaborate more on this in Section V.

2) Geometry and data-based computation. The gap metric has a well-known geometric interpretation in terms of the sine of the largest principal angle between two subspaces [32]. In particular, as an immediate consequence of [32, Theorem 4.5], for $L > l$, the L-gap between $\mathcal{B} \in \mathcal{L}^{q,c}$ and $\tilde{\mathcal{B}} \in \mathcal{L}^{q,c}$ is $\text{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) = \sin \theta_{\max}$, where $\theta_{\max}$ is the largest principal angle between the subspaces $\mathcal{B}|_L$ and $\tilde{\mathcal{B}}|_L$. Furthermore, by Proposition 1 and since gap is a metric on $\text{Gr}(k, N)$ for $k, N \in \mathbb{N}$ [32, p.93], we have the following result.

Corollary 1. The set of all restricted behaviors $\mathcal{B}|_L \in \mathcal{L}^{q,c}$, with $L > l$, equipped with gap_L is a metric space.

The L-gap between behaviors can be directly computed from the knowledge of sufficiently excited trajectories. Let $w^d \in \mathcal{B}|_T$ and $\tilde{w}^d \in \tilde{\mathcal{B}}|_T$ be sufficiently excited T-length trajectories of order L, with $L > l$. Let

$$H_L(w^d) = [\, U_1 \, U_2 \,] \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix},$$

$$H_L(\tilde{w}^d) = [\, \tilde{U}_1 \, \tilde{U}_2 \,] \begin{bmatrix} \tilde{S} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{V}_1 \\ \tilde{V}_2 \end{bmatrix}$$

be the singular value decomposition (SVD) of the Hankel matrices $H_L(w^d)$ and $H_L(\tilde{w}^d)$ with $U_1 \in \mathbb{R}^{qL \times (mL+n)}$ and $\tilde{U}_1 \in \mathbb{R}^{qL \times (mL+n)}$, respectively. Then

$$\begin{aligned} \text{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) &= \|H_L(w^d)H_L(w^d)^\dagger - H_L(\tilde{w}^d)H_L(\tilde{w}^d)^\dagger\|_2 \\ &= \|U_1 U_1^\mathsf{T} - \tilde{U}_1 \tilde{U}_1^\mathsf{T}\|_2 = \|\tilde{U}_2^\mathsf{T} U_1\|_2 \end{aligned} \tag{6}$$

where the first identity follows from $\text{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) = \|P_{\mathcal{B}|_L} - P_{\tilde{\mathcal{B}}|_L}\|_2$ and since $P_{\mathcal{B}|_L} = U_1 U_1^\mathsf{T}$ and $P_{\tilde{\mathcal{B}}|_L} = \tilde{U}_1 \tilde{U}_1^\mathsf{T}$ [34, p.82]. The second identity is due to [34, Thm 2.5.1].

B. Uncertainty quantification

The following theorem, which is inspired by [35, Prop. 7], provides an upper and a lower bound on the L-gap in case the restricted behaviors have a specific form.

Theorem 1. Let $\mathcal{B} \in \mathcal{L}^{q,c}$ and $\tilde{\mathcal{B}} \in \mathcal{L}^{q,c}$. Given $L \in \mathbb{Z}_+$, with $L > l$, assume

$$\mathcal{B}|_L = \mathrm{im} \begin{bmatrix} I \\ F \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{B}}|_L = \mathrm{im} \begin{bmatrix} I \\ \tilde{F} \end{bmatrix}.$$

Then

$$\frac{\|F - \tilde{F}\|_2}{\sqrt{1 + \|F\|_2^2}\sqrt{1 + \|\tilde{F}\|_2^2}} \leq \mathrm{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) \leq \|F - \tilde{F}\|_2.$$

Theorem 1 can be used to quantify uncertainty for a wide class of systems. For example, assume $\mathcal{B} \in \mathcal{L}^{q,c}$ is defined by the AR model

$$y_{t+L-1} = \sum_{k=0}^{L-2} a_k y_{t+k} + \sum_{k=0}^{L-1} b_k u_{t+k}, \tag{7}$$

where $a_k \in \mathbb{R}$, $b_k \in \mathbb{R}$, and $L \in \mathbb{N}$ is such that $L > l$. Then

$$\mathcal{B}|_L = \ker \begin{bmatrix} F & -I \end{bmatrix} = \mathrm{im} \begin{bmatrix} I \\ F \end{bmatrix},$$

with $F = \begin{bmatrix} a_0 & b_0 & \dots & a_{L-2} & b_{L-2} & b_{L-1} \end{bmatrix}$. Based on this observation, we have the following corollary for AR models.

Corollary 2. Let $\mathcal{B} \in \mathcal{L}^{q,c}$ and $\tilde{\mathcal{B}} \in \mathcal{L}^{q,c}$. Given $L \in \mathbb{Z}_+$, with $L > l$, assume $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are defined by the AR models (7) and

$$\tilde{y}_{t+L-1} = \sum_{k=0}^{L-2} \tilde{a}_k \tilde{y}_{t+k} + \sum_{k=0}^{L-1} \tilde{b}_k \tilde{u}_{t+k}, \tag{8}$$

with $\tilde{a}_k \in \mathbb{R}$ and $\tilde{b}_k \in \mathbb{R}$, respectively. Assume

$$F = \begin{bmatrix} a_0 & b_0 & \dots & a_{L-2} & b_{L-2} & b_{L-1} \end{bmatrix},$$
$$\tilde{F} = \begin{bmatrix} \tilde{a}_0 & \tilde{b}_0 & \dots & \tilde{a}_{L-2} & \tilde{b}_{L-2} & \tilde{b}_{L-1} \end{bmatrix}.$$

are such that $\|F - \tilde{F}\|_2 \leq \epsilon$. Then $\mathrm{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) \leq \epsilon$.

Remark 1. We presented Corollary 2 for the AR models (7) and (8) for clarity of exposition, but the result holds for more general multi-input multi-output systems. Corollary 2 raises the natural open question of how to relate the L-gap to classical uncertainty models [18–21] including additive, multiplicative, and coprime factor uncertainties. This is left as an area of future work. $\triangle$

C. Connection with the gap metric on ℓ_2

The gap metric plays a central role in robust control theory [18–21], where finite-dimensional, LTI systems are regarded as operators acting on a given Hilbert space $\mathcal{H}$, such as[1] ℓ_2 or H_2. In this context, the distance between finite-dimensional, LTI systems is defined in terms of the

gap between the graphs of the corresponding input-output operators.

Definition 9. [19, p.30] Let[2] $\mathcal{H} = \mathcal{U} \times \mathcal{Y}$, with $\mathcal{U}$ and $\mathcal{Y}$ Hilbert spaces. Let $P : \mathcal{V} \to \mathcal{Y}$ and $\tilde{P} : \tilde{\mathcal{V}} \to \mathcal{Y}$ be closed operators, with $\mathcal{V}$ and $\tilde{\mathcal{V}}$ subspaces of $\mathcal{U}$. The *gap between* P *and* $\tilde{P}$ is defined as

$$\mathrm{gap}_{\mathcal{H}}(P, \tilde{P}) = \mathrm{gap}_{\mathcal{H}}(\mathrm{graph}(P), \mathrm{graph}(\tilde{P})).$$

The gap metric on H_2 admits a classical interpretation in terms of Nyquist diagrams (plotted on the Riemann sphere) and can be computed as the H_∞ norm of a transfer matrix [21]. We now establish that, under certain assumptions, the L-gap metric can be connected to the gap metric on ℓ_2.

Theorem 2. Let $\mathcal{B} \in \mathcal{L}^{q,c}$ and $\tilde{\mathcal{B}} \in \mathcal{L}^{q,c}$. Assume $\mathcal{B} = \mathrm{graph}(P)$ and $\mathcal{B} = \mathrm{graph}(\tilde{P})$, with P and $\tilde{P}$ bounded linear operators on ℓ_2. Then

$$\lim_{L \to \infty} \mathrm{gap}_L(\mathcal{B}, \tilde{\mathcal{B}}) \leq \mathrm{gap}_{\ell_2}(P, \tilde{P}).$$

IV. APPLICATION: MODE RECOGNITION AND CONTROL

We envision many applications of the L-gap metric, e.g., prediction error quantification, robustification in data-driven control, and fault detection and isolation. In the following case study, we use it as an analysis tool in the spirit of mode recognition and control. Namely, we determine the mode of a switched autoregressive exogenous (SARX) system directly from data for the purpose of data-driven control.

Consider a SARX system [36] with 2 modes given by

$$\begin{aligned} y_t &= 0.2 y_{t-1} + 0.24 y_{t-2} + 2 u_{t-1} + n_t, \\ y_t &= 0.7 y_{t-1} - 0.12 y_{t-2} + 1 u_{t-1} + n_t, \end{aligned} \tag{9}$$

where $u_t \in \mathbb{R}$ and $y_t \in \mathbb{R}$ are the inputs and outputs at time $t \in \mathbb{Z}_+$, and $n_t \sim \mathcal{N}(0, \sigma^2)$ is observation noise with $\sigma = 10^{-4}$ and truncated to the interval $[-3\sigma, 3\sigma]$.

We consider the problem of performing data-driven control [6], while recognizing switches in the system's mode. To this end, we use DeePC [6] which solves the following optimal control problem in a receding horizon fashion for some data matrix D serving as a predictive model for allowable trajectories:

$$\begin{aligned} \underset{y,u,g}{\text{minimize}} \quad & 2000\|y - r\|^2 + \|u\|^2 + 20\|g\|^2 \\ \text{subject to} \quad & Dg = (u_{\mathrm{ini}}, u, y_{\mathrm{ini}}, y), \end{aligned} \tag{10}$$

where $(u_{\mathrm{ini}}, y_{\mathrm{ini}})$ is the most recent T_{ini}-length trajectory of the system (used to implicitly fix the initial condition from which the T_{f}-length prediction, (u, y) evolves), and $r \in \mathbb{R}^{T_{\mathrm{f}}}$ is a given reference trajectory. We select $T_{\mathrm{ini}} = 2$ and $T_{\mathrm{f}} = 5$.

By Lemma 3, any data matrix D containing sufficiently exciting data from a particular system mode describes (approximately due to noise) the subspace in which trajectories live for that mode. By performing an SVD of D, we can identify a large decrease in the singular values indicating the dimension of the subspace of allowable trajectories. Note that SVD

---

[1] ℓ_2 is the Hilbert space of square summable sequences $u : \mathbb{Z}_+ \to \mathbb{R}^m$. H_2 is the Hardy space of functions $f : \mathbb{C} \to \mathbb{C}^m$ which are analytic in the complement of the closed unit disk [19, p.13].

[2] The *graph* of an operator $P : \mathcal{U} \to \mathcal{Y}$ is defined as $\mathrm{graph}(P) = \{(u, Pu) \in \mathcal{U} \times \mathcal{Y} : u \in \mathcal{U}\}$ [19, p.17].

does not necessarily preserve structure, but we only require a basis for the restricted behavior. In this case, the subspace dimension is given by (1) and is equal to $T_{\text{ini}} + T_{\text{f}} + n = 9$, as shown in Fig. 1.
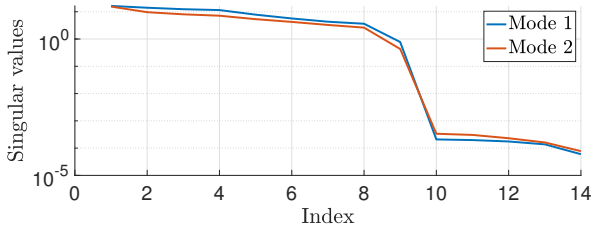

Fig. 1: Singular values of data matrices representing restricted behaviors of each mode of system (9).

Distinguishing the modes would not be possible by looking at the singular values of the data matrices alone. We propose the following data-driven mode recognition and control strategy based on the use of the L-gap to distinguish the modes. Before starting, let $t \geq 0$ denote the current time, and fix the matrix D in (10). The first step is to compute an SVD of D and form a basis D_{basis} using the first $T_{\text{ini}} + T_{\text{f}} + n$ left singular vectors. Next compute an SVD of a matrix with M columns containing the most recent $T_{\text{ini}} + T_{\text{f}}$-length trajectories, denoted H_t, and form a basis, denoted $H_{t,\text{basis}}$, using the first $T_{\text{ini}} + T_{\text{f}} + n$ left singular vectors. Fix a threshold $\epsilon > 0$. If $\text{gap}_{T_{\text{ini}} + T_{\text{f}}}(\text{im } H_{t,\text{basis}}, \text{im } D_{\text{basis}}) > \epsilon$, set $D = H_t$. This can be thought of as adopting the most recent data as the predictive model in (10) only when the gap between the predictive model D and the most recent data H_t is larger than some pre-defined threshold. Equipped with the data matrix D, solve (10) for the optimal predicted input trajectory $(u_1^\star, \ldots, u_{T_{\text{f}}}^\star)$ and apply $u_t = u_1^\star$ to the system. Measure y_t and set $(u_{\text{ini}}, y_{\text{ini}})$ in (10) to the most recent T_{ini}-length trajectory of the system. Update H_t by deleting the first column and adding the most recent $T_{\text{ini}} + T_{\text{f}}$-length trajectory as the last column. This process is repeated to perform simultaneous data-driven mode recognition and control.

The strategy above has been simulated with $\epsilon = 0.3$ on system (9) for $t \in [0, 70]$. We arbitrarily initialize the predictive model D in (10) to be a matrix containing sufficiently exciting data from mode 1. However, the system starts in mode 2 and only switches to mode 1 at $t = 40$. The strategy is compared to data-driven control without mode recognition, i.e., where D is kept constant in (10). The results are shown in Fig. 2 and Fig. 3. We observe in Figure 2 that the controlled output trajectory is offset from the desired reference. This is due to the fact that we are using the wrong data set in (10) for predicting optimal trajectories. However, the L-gap between D_{basis} and $H_{t,\text{basis}}$ quickly increases above the threshold ϵ, thus successfully recognizing a discrepancy between the current mode of the system and the data being used for control (see Fig. 3). During this transient phase, the moving window contains a mixture of data containing trajectories from mode 2, and mode 1. However, at approximately $t = 22$, the L-gap successfully
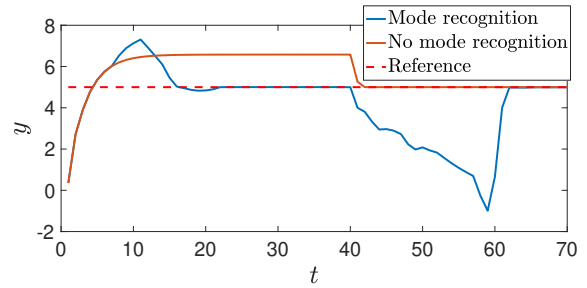

Fig. 2: Performance of mode recognition and control strategy on the SARX system with switches between modes compared to data-driven control without mode recognition.

recognizes that the data matrix D used in (10) is consistent with the current mode of the system (mode 2). The control performance after this transient phase then improves. This is again illustrated during the mode switch at $t = 40$. On the other hand, the data-driven control strategy with no mode recognition does not adapt to mode switches and has poor performance until $t = 40$ where the system switches incidentally to mode 1 thus matching with the fixed data matrix D being used in this strategy. This case study suggests that the L-gap is a suitable tool for data-driven online mode recognition and control.
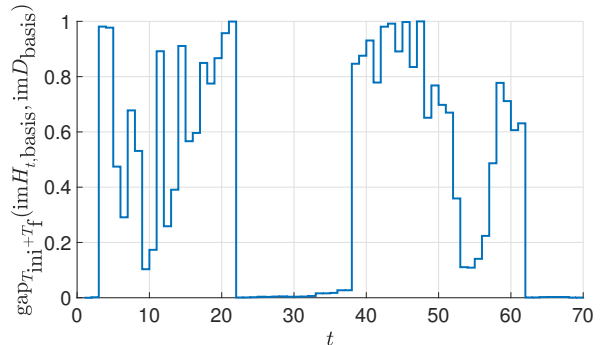

Fig. 3: Distance computed with the L-gap between $H_{t,\text{basis}}$ and D_{basis} for $L = T_{\text{ini}} + T_{\text{f}}$.

V. Conclusion

This paper has explored the issue of uncertainty quantification in the behavioral setting. A new metric has been defined on the set of restricted behaviors and shown to capture parametric uncertainty for the class of AR models. The metric is a direct finite-time counterpart of the classical gap metric. A data-driven control case study has illustrated the value of the new metric through numerical simulations.

The paper has shown that the gap induces a metric space structure on the set of restricted behaviors. However, there are many other common metrics defined on Grassmannians [37]. In fact, all such distances depend on the principal angles. This is not a coincidence. The geometry of the Grassmannian is such that any rotationally invariant metric between k-dimensional subspaces in $\mathbb{R}^N$ (i.e., dependent only on the

relative the position of subspaces) is necessarily a function of the principal angles.

Theorem 3. [33, Theorem 2] Let d be a rotationally invariant[3] metric on $\mathrm{Gr}(k, N)$. Then $d(\mathcal{V}, \mathcal{W})$ is a function of the principal angles between the subspaces $\mathcal{V}$ and $\mathcal{W}$.

Each metric induces a particular geometry, which comes with its own advantages and disadvantages. For example, the *Grassmann metric* is the geodesic distance on $\mathrm{Gr}(k, N)$ [33, Theorem 2], viewed as a Riemannian (quotient) manifold and the corresponding geodesics admit an explicit expression [38]. This, in turn, suggests that the choice of the metric structure on the set of restricted behaviors is crucial, raising a number of important questions. For instance, any metric on $\mathrm{Gr}(k, N)$ that induces a differentiable structure opens up the possibility of directly optimizing over behaviors. So can one exploit any such structure to improve the performance of data-driven control algorithms (e.g., DeePC)? In practice, non-parametric representations of restricted behaviors are typically constructed from noisy measurements. This can be a serious drawback, because noisy restricted behaviors may appear to be far apart, even when close and/or of the same dimension. This issue may be resolved by considering metrics on infinite Grassmannians [33]. So can one leverage these results in an online, real-time, noisy setting where behaviors are constantly changing? We leave the exploration of these important questions as future research directions.

REFERENCES

[1] L. Ljung, *System identification - Theory for the user (2nd edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[2] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: theory - implementation - applications*. Dordrecht, The Netherlands: Kluwer, 1996.

[3] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor, "A note on persistency of excitation," *Syst. Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.

[4] J. C. Willems, "From time series to linear system—Part I. Finite dimensional linear time invariant systems," *Automatica*, vol. 22, no. 5, pp. 561–580, 1986.

[5] J. C. Willems and J. W. Polderman, *Introduction to mathematical systems theory: a behavioral approach*. New York, NY, USA: Springer, 1997.

[6] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the deepc," in *Proc. Eur. Control Conf.*, Naples, Italy, 2019, pp. 307–312.

[7] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 909–924, 2019.

[8] H. J. van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel, "Data informativity: a new perspective on data-driven analysis and control," *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4753–4768, 2020.

[9] I. Markovsky and F. Dörfler, "Behavioral systems theory in data-driven analysis, signal processing, and control," *Ann. Rev. Control*, vol. 52, pp. 42–64, 2021.

[10] P. G. Carlet, A. Favato, S. Bolognani, and F. Dörfler, "Data-driven predictive current control for synchronous motor drives," in *IEEE Energy Conversion Congress and Exposition*, Detroit, MI, USA, 2020, pp. 5148–5154.

[11] L. Huang, J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control for grid-connected power converters," in *Proc. 58th Conf. Decision Control*, Naples, Italy, 2019, pp. 8130–8135.

[12] J. Coulson, J. Lygeros, and F. Dörfler, "Distributionally robust chance constrained data-enabled predictive control," *IEEE Trans. Autom. Control*, 2021.

[13] J. Berberich, A. Koch, C. W. Scherer, and F. Allgöwer, "Robust data-driven state-feedback design," in *Proc. Amer. Control Conf.*, Denver, CO, USA, 2020, pp. 1532–1538.

[14] H. J. van Waarde, M. K. Camlibel, and M. Mesbahi, "From noisy data to feedback controllers: Nonconservative design via a matrix S-lemma," *IEEE Transactions on Automatic Control*, vol. 67, no. 1, pp. 162–175, 2022.

[15] A. Xue and N. Matni, "Data-driven system level synthesis," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 189–200.

[16] A. Bisoffi, C. De Persis, and P. Tesi, "Data-driven control via Petersen's lemma," *arXiv:2109.12175*, 2021.

[17] L. Huang, J. Zhen, J. Lygeros, and F. Dörfler, "Robust data-enabled predictive control: Tractable formulations and performance guarantees," *arXiv preprint arXiv:2105.07199*, 2021.

[18] G. Zames and A. El-Sakkary, "Uncertainty in unstable systems: The gap metric," in *Proc. 8th IFAC World Congr.*, vol. 14, no. 2, 1981, pp. 149–152.

[19] J. R. Partington, *Linear operators and linear systems: an analytical approach to control theory*. Cambridge, U. K.: Cambridge Univ. Press, 2004.

[20] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. New Jersey, NJ, USA: Prentice Hall, 1996.

[21] G. Vinnicombe, *Uncertainty and feedback: H_∞ loop-shaping and the ν-gap metric*. London, U.K.: Imperial College Press, 2001.

[22] A. J. Sasane, "Distance between behaviours," *Int. J. Control*, vol. 76, no. 12, pp. 1214–1223, 2003.

[23] J. A. Ball and A. J. Sasane, "Equivalence of a behavioral distance and the gap metric," *Syst. Control Lett.*, vol. 55, no. 3, pp. 214–222, 2006.

[24] H. L. Trentelman and S. V. Gottimukkala, "Distance between behaviors and rational representations," *SIAM J. Control Optim.*, vol. 51, no. 6, pp. 4211–4241, 2013.

[25] A. Padoan, J. Coulson, H. J. van Waarde, J. Lygeros, and F. Dörfler, "Behavioral uncertainty quantification for data-driven control," 2022, arXiv preprint, arXiv:2204.02671, April 2022.

[26] F. Dorfler, J. Coulson, and I. Markovsky, "Bridging direct & indirect data-driven control formulations via regularizations and relaxations," *IEEE Trans. Autom. Control*, 2022.

[27] K. J. Åström and B. Wittenmark, *Adaptive control (2nd edition)*. Reading, MA, USA: Addison-Wesley, 1995.

[28] I. Markovsky and F. Dörfler, "Identifiability in the behavioral setting," *Vrije Universiteit Brussel, Tech. Rep*, 2020. [Online]. Available: http://homepages.vub.ac.be/~imarkovs/publications/identifiability.pdf.

[29] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry (2nd Edition)*. London, U. K.: Academic Press, 1986.

[30] N. L. Carothers, *Real analysis*. Cambridge, U. K.: Cambridge Univ. Press, 2000.

[31] T. Kato, *Perturbation theory for linear operators (2nd Edition)*. New York, NY, USA: Springer, 1980.

[32] G. Stewart and J. Sun, *Matrix Perturbation Theory*. New York, NY, USA: Academic Press, 1990.

[33] K. Ye and L. H. Lim, "Schubert varieties and distances between subspaces of different dimensions," *SIAM J. Mat. Anal. Appl.*, vol. 37, no. 3, pp. 1176–1197, 2016.

[34] G. H. Golub and C. F. Van Loan, *Matrix computations (4th Ed.)*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.

[35] L. Qui and E. J. Davison, "Pointwise gap metrics on transfer matrices," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 741–758, 1992.

[36] Z. Du, L. Balzano, and N. Ozay, "A robust algorithm for online switched system identification," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 293–298, 2018.

[37] M. M. Deza and E. Deza, *Encyclopedia of distances*. New York, NY, USA: Springer, 2009.

[38] P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

---

[3]A metric d on $\mathrm{Gr}(k, N)$ is *rotationally invariant* if

$$d(Q \cdot \mathcal{V}, Q \cdot \mathcal{W}) = d(\mathcal{V}, \mathcal{W})$$

for all $Q \in O(N)$ and $\mathcal{V}, \mathcal{W} \in \mathrm{Gr}(k, N)$ [33, p.1179], where the *left action* of the orthogonal group $O(N)$ on $\mathrm{Gr}(k, N)$ is defined as $Q \cdot \mathcal{V} = \mathrm{im}(QV)$ for $Q \in O(N)$ and $\mathcal{V} \in \mathrm{Gr}(k, N)$.