Convex Behavioral Model (In)Validation via Jensen-Bregman Divergence Minimization

O. Camps M. Sznaier X. Zhang
Electrical and Comp. Engineering Department,
Northeastern University,
Boston, MA 02115.

***Abstract*— This paper considers the problem of determining whether two pairs of (noisy) time sequences $(\mathbf{u}, \mathbf{y})$ are behaviors of the same (unknown) underlying system. That is, whether these pairs are admissible input/output trajectories for some suitable initial condition. This problem is relevant to many practical scenarios arising not only in the context of control and systems identification, e.g. model (in)validation and fault detection, but also in other fields, including computer vision and image processing. Our main result shows that this problem can be reduced to minimizing the Jensen-Bregman log-det divergence between two suitable constructed Grammian matrices, a problem that can be efficiently solved using recently introduced methods. This result is illustrated with two non–trivial examples: activity recognition from video sequences and fault detection in a bearing rig.**

I. Introduction and Motivation

A large number of practical scenarios require determining whether two given (noisy) pairs of time series can be considered to be input/output trajectories of the same system, for some unknown initial conditions. For instance, this problem is a pre-requisite to data-driven identification of piece-wise affine models, where usually the first step is to determine regions where the experimental data can be explained by a single model. In addition, this problem also arises in the context of fault detection, since typically faults cause a change in the underlying dynamics and thus a single system cannot explain the observed data record. A similar reasoning can be used in video-analytics to detect anomalies from video sequences. Finally, by postulating that activities are manifestations of the same underlying dynamics, this approach can also be used for activity classification (see for instance [2] and references therein).

Formally, the problem above can be stated as a behavioral model (in)validation problem and solved using tools developed in this context. For instance, it is possible to use a two-step approach based on (i) first finding the most powerful unfalsified model [3] that explains one of the sequences and (ii) establishing whether this model admits the second sequence as a behavior. While this approach works well with clean data, it may fail in the presence of measurement noise. Noisy sequences and model uncertainty can be handled by pursuing the rank-minimization approach proposed in [8]. Since rank minimization is known to be NP-hard, a convex relaxation, based on using the nuclear norm as a surrogate for rank, is used instead. Thus, there is no guarantee that this approach will, in all cases, correctly label the behaviors.

To avoid these difficulties, in this paper we propose an (in)validation approach based on computing the Jensen-Bregman log det divergence between Gram matrices built from the experimental data. As in [8], the paper is based on earlier results from subspace identification theory [4], relating the input/ouput Hankel matrices of an LTI system. However, rather than using these results to directly compare ranks, as in [8], in this paper we show that two behaviors originate from the same system if and only if the distance (in the Jensen-Bregman log det sense) between suitable regularized Gram matrices is finite, and relate the distance between trajectories to the distance between initial conditions. This leads to an (in)validation algorithm based upon searching for a positive semi-definite matrix (related to the noise) that minimizes the distance between the "denoised" trajectories. An advantage of this approach is that this minimization problem is convex in cases where the distance between initial conditions is not too large (in a sense precisely defined in the body of the paper). Further, even in cases where this condition fails, efficient solution methods exist based on "convex plus concave" decompositions [9]. The paper finishes by illustrating these results with an academic example and two realistic problems: activity recognition and anomaly detection.

II. Preliminaries

For ease of reference, we summarize next the notation used in the paper and recall some results required to recast the invalidation problem into a convex optimization form.

A. Notation

$\mathbb{R}$	set of real numbers
$\mathbf{x}, \mathbf{M}$	a vector in $\mathbb{R}^n$ (matrix in $\mathbb{R}^{n\times m}$)
$\mathbf{M}^T$	transpose of matrix $\mathbf{M}$.
$\overline{\sigma}_{\mathbf{M}}$	maximum singular value of $\mathbf{M}$.
$\sigma_{\mathbf{M},i}$	i^{th} singular value of $\mathbf{M}$.
$\lvert\mathbf{M}\rvert$	determinant of $\mathbf{M}$.
$\mathbf{M}^{\perp}$	(right) annihilator of $\mathbf{M}$: $\mathbf{M}\mathbf{M}^{\perp} = \mathbf{0}$ and $(\mathbf{M}^{\perp})^T\mathbf{M}^{\perp} = \mathbf{I}$

This work was supported in part by NSF grants IIS–1318145 and ECCS–1404163; AFOSR grant FA9550-15-1-0392; and the Alert DHS Center of Excellence under Award Number 2013-ST-061-ED0001. email {camps,msznaier,zhangxk}@ece.neu.edu

$\mathbf{M} \succeq \mathbf{N}$	the matrix $\mathbf{M}-\mathbf{N}$ is positive semidefinite.
$\mathcal{S}^n$	set of symmetric matrices in $\mathbb{R}^{n\times n}$
$\mathcal{S}^n_+(\mathcal{S}^n_{++})$	cone of positive-semidefinite (-definite) matrices in $\mathcal{S}^n$
$\dim(X)$	dimension of the subspace X
$\mathcal{E}(\mathbf{X})$	expected value of a stochastic matrix $\mathbf{X}$.
$J_{ld}(\mathbf{X},\mathbf{Y})$	Jensen-Bregman Log Det Divergence:

$$J_{ld}(\mathbf{X},\mathbf{Y}) = \log\left|\frac{\mathbf{X}+\mathbf{Y}}{2}\right| - \frac{1}{2}\log|\mathbf{X}\mathbf{Y}|,\ \mathbf{X},\mathbf{Y}\in\mathcal{S}^n_{++}$$

$\mathbf{H}^{m,n}_{\mathbf{y}}(\mathbf{T}^n_{\mathbf{y}})$	Hankel (Toeplitz) matrix associated with a vector sequence $\mathbf{y}(.)$:

$$\mathbf{H}^{m,n}_y \doteq \begin{bmatrix} \mathbf{y}(0) & \mathbf{y}(1)\cdots & \mathbf{y}(m) \\ \mathbf{y}(1) & \mathbf{y}(2)\cdots & \mathbf{y}(m+1) \\ \vdots & \vdots\ddots & \vdots \\ \mathbf{y}(n) & \mathbf{y}(n+1)\cdots & \mathbf{y}(m+n-1) \end{bmatrix}$$

$$\mathbf{T}^n_y \doteq \begin{bmatrix} \mathbf{y}(0) & \mathbf{0}\cdots & \mathbf{0} \\ \mathbf{y}(1) & \mathbf{y}(0)\cdots & \mathbf{0} \\ \vdots & \vdots\ddots & \vdots \\ \mathbf{y}(n-1) & \mathbf{y}(n-2)\cdots & \mathbf{y}(0) \end{bmatrix}$$

$\mathbf{\Gamma}_{\mathbf{A},\mathbf{C}}$	Observability matrix associated with the pair $(\mathbf{A},\mathbf{C})$.

$$\mathbf{\Gamma}^T_{\mathbf{A},\mathbf{C}} \doteq \begin{bmatrix}\mathbf{C}^T\ (\mathbf{C}\mathbf{A})^T\ \ldots\ (\mathbf{C}\mathbf{A}^i)^T\ \ldots\end{bmatrix}^T$$

For simplicity, the dimensions of $\mathbf{H}$ and $\mathbf{T}$ may be omitted, when clear from the context. Further, by a slight abuse of notation, given a strictly proper plant G with state space representation:

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{A}\mathbf{x}(k)+\mathbf{B}\mathbf{u}(k), \\ \mathbf{z}(k) &= \mathbf{C}\mathbf{x}(k), \\ \mathbf{y}(k) &= \mathbf{z}(k)+\mathbf{v}(k) \end{aligned} \tag{1}$$

we will denote by $\mathbf{H}_G$ and $\mathbf{T}_G$ the Hankel and Toeplitz matrices associated with its impulse response sequence $\mathbf{z}(k) = \mathbf{C}\mathbf{A}^{\mathbf{k}}\mathbf{B},\ k=0,1,\ldots$.

B. Problem Statement

In this paper we consider the problem of determining whether two given (noisy) input/ouput pairs $(\mathbf{u}_1,\mathbf{y}_1)$ and $(\mathbf{u}_2,\mathbf{y}_2)$ are behaviors of the same (unknown) LTI system. That is, whether $\mathbf{z}_1 \doteq \mathbf{y}_1-\mathbf{v}_1$ and $\mathbf{z}_2 \doteq \mathbf{y}_2-\mathbf{v}_2$ are solutions of a set of equations of the form (1) for some triple $(\mathbf{A},\mathbf{B},\mathbf{C})$ and suitable initial conditions and noise sequences $\mathbf{v}_1,\mathbf{v}_2$. Formally, this leads to the following problem:

Problem 1: Given two input/output pairs $(\mathbf{u}_1(k),\mathbf{y}_1(k))$, $k=t_1,\ldots,t_{T_1}$ and $(\mathbf{u}_2(k),\mathbf{y}_2(k))$, $k=t_2,\ldots,t_{T_2}$, and a bound η on the covariance of the measurement noise, determine whether there exists a triple $(\mathbf{A},\mathbf{B},\mathbf{C})$, initial conditions $\mathbf{x}_{0,1}$, $\mathbf{x}_{0,2}$ and admissible noise sequences $\{\mathbf{v}_1(k)\},\{\mathbf{v}_2(k)\}$, $\|\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^T)\|_2\le\eta$ such that (1) holds with $\mathbf{z}_i(k)\doteq\mathbf{y}_i(k)-\mathbf{v}_i(k)$

In the sequel, in order to obtain deterministic (in)validation certificates, proceeding as in [6] we will use deterministic descriptions of the measurement noise $\mathbf{v}$. Thus, we will replace the constraint $\mathcal{E}(\mathbf{v}\mathbf{v}^T)\le\eta$ with

$$\|\mathbf{G}_{\mathbf{v}}\|_2 \le \epsilon \doteq \frac{1}{n}\eta \tag{2}$$

where $\mathbf{G}_{\mathbf{v}} \doteq \mathbf{H}^n_{\mathbf{v}}(\mathbf{H}^n_{\mathbf{v}})^T$.

C. Background Results

In this section we review some results that will allow for recasting Problem 1 into a convex minimization form.

Lemma 1: [4] Assume that the pairs $(\mathbf{A},\mathbf{B})$ and $(\mathbf{A},\mathbf{C})$ are controllable and observable, respectively, with $\mathbf{A}\in\mathbb{R}^{n\times n}$. Then:

$$\mathbf{H}_{\mathbf{y}} = \mathbf{\Gamma}_{\mathbf{A},\mathbf{C}}\mathbf{X} + \mathbf{T}_G\mathbf{H}_{\mathbf{u}} + \mathbf{H}_{\mathbf{v}} \tag{3}$$

where $\mathbf{X}\doteq\begin{bmatrix}\mathbf{x}(\mathbf{0})\ \mathbf{x}(\mathbf{1})\ \ldots\ \mathbf{x}(\mathbf{t})\end{bmatrix}$. Further, if

$$span_{row}(\mathbf{X})\cap span_{row}(\mathbf{H}_{\mathbf{u}}) = \{0\} \tag{4}$$

then

$$\text{rank}(\mathbf{H}_{\mathbf{y}}\mathbf{H}_{\mathbf{u}}{}^{\perp}) = \text{rank}(\mathbf{X})$$

In the sequel, in order to handle noisy sequences, we will replace condition (4) with the slightly stronger one:

$$span_{row}(\mathbf{X}) \perp span_{row}(\mathbf{H}_{\mathbf{u}}) \tag{5}$$

(or, equivalently, $\mathbf{X}\cdot\mathbf{U}=0$ where $\mathbf{U}$ is an orthonormal basis of the row space of $\mathbf{H}_{\mathbf{u}}$[1]).

Lemma 2: [1] For a fixed $\mathbf{Y}\in\mathcal{S}^n_{++}$, $J_{ld}(\mathbf{X},\mathbf{Y})$ is convex in the region $\{\mathbf{X}\in\mathcal{S}^n_{++}:\mathbf{X}\preceq(1+\sqrt{2})\mathbf{Y}\}$.

III. Main Results

In this section we show that Problem 1 can be recast into a J_{ld} minimization form. To establish this result we will first analyze the behavior of the Jensen-Bregman divergence between suitably regularized positive semi-definite matrices and then combine these results with Lemmas 1 and 2 to obtain computationally tractable (in)validation certificates. Due to space constraints most proofs have been omitted. They can be obtained by contacting the authors to request an extended version of the paper.

A. Jensen-Bregman divergence between Grammians

The first step towards obtaining the (in)validation certificates proposed in this paper is to show that the J_{ld} distance between two suitable matrices $\mathbf{X},\mathbf{Y}\in\mathcal{S}^n_+$ is finite if and only if these matrices share the same null-space.

Lemma 3: Let $\mathbf{X},\mathbf{Y}\in\mathcal{S}^n_+$ and define $\hat{\mathbf{X}}(\sigma)\doteq\mathbf{X}+\sigma\mathbf{I}$ and $\hat{\mathbf{Y}}(\sigma)\doteq\mathbf{Y}+\sigma\mathbf{I}$. Assume, without loss of generality, that $rank(\mathbf{X})=r_{\mathbf{X}}\ge rank(\mathbf{Y})=r_{\mathbf{Y}}$. Then $\lim_{\sigma\to0}J_{ld}(\frac{1}{2}(\hat{\mathbf{X}}+\hat{\mathbf{Y}}),\hat{\mathbf{X}})<\infty \iff span_{col}(\mathbf{X}^{\perp})\subseteq span_{col}(\mathbf{Y}^{\perp})$.

[1] In a stochastic framework this condition simply imposes that the state at a time k is uncorrelated with future values of the input.

Next, we exploit the result above to establish that two given input/output pairs originate from systems sharing the same $(\mathbf{A}, \mathbf{C})$ if the J_{ld} distance between the corresponding (regularized) Grammians is finite. Further, in this case, this distance is a measure of the distance between the orthogonal projection of the state-space trajectories on the null space of $\mathbf{H}_u^{\perp}$, that is, roughly speaking, the portion of the trajectories that is due to the effect of initial conditions.

Theorem 1: Given two input/output pairs $(\mathbf{u}_i, \mathbf{z}_i),\ i = 1, 2$, define the matrices

$$\mathbf{G}_{\mathbf{z}i} \doteq (\mathbf{H}_{\mathbf{z}i}\mathbf{H}_{\mathbf{u}i}^{\perp})(\mathbf{H}_{\mathbf{z}i}\mathbf{H}_{\mathbf{u}i}^{\perp})^T \tag{6}$$

Assume that $rank(\mathbf{G}_{\mathbf{z}1}) = rank(\mathbf{G}_{\mathbf{z}2}) = n$ and condition (4) holds. Then,

(i) There exists a pair $(\mathbf{A}, \mathbf{C})$ and initial conditions $\mathbf{x_{1,o}}$ and $\mathbf{x_{2,o}}$ such that (1) is satisfied iff

$$\lim_{\sigma\to 0} J_{ld}\left(\frac{1}{2}(\mathbf{G}_{\mathbf{z}1} + \mathbf{G}_{\mathbf{z}2}) + \sigma\mathbf{I}, \mathbf{G}_{\mathbf{z}1} + \sigma\mathbf{I}\right) < \infty \tag{7}$$

(ii) If (7) holds, then

$$\begin{aligned}&\lim_{\sigma\to 0} J_{ld}\left(\tfrac{1}{2}(\mathbf{G}_{\mathbf{z}1} + \mathbf{G}_{\mathbf{z}2}) + \sigma\mathbf{I}, \mathbf{G}_{\mathbf{z}1} + \sigma\mathbf{I}\right) = \\ &J_{ld}(\tfrac{1}{2}(\mathbf{X_1}\mathbf{H_{u}}_1^{\perp}(\mathbf{H_u}_1^{\perp})^T\mathbf{X}_1^T) + \\ &\qquad \tfrac{1}{2}(\mathbf{X_2}\mathbf{H_u}_2^{\perp}(\mathbf{H_u}_2^{\perp})^T\mathbf{X}_2^T), \\ &\qquad\qquad \mathbf{X_1}\mathbf{H_u}_1^{\perp}(\mathbf{H_u}_1^{\perp})^T\mathbf{X}_1^T)\end{aligned} \tag{8}$$

where $\mathbf{X}_1, \mathbf{X}_2$ denote the corresponding state trajectories[2].

B. Handling noisy sequences

In this section we consider the effects of measurement noise and show that, as long as this noise is uncorrelated with the state trajectories, the results of the previous section can be used to recast Problem 1 into an optimization over matrices in $\mathcal{S}_+^n$.

Theorem 2: Given two input/output pairs $(\mathbf{u}_i, \mathbf{y}_i),\ i = 1, 2$, define the matrices

$$\mathbf{G}_{\mathbf{y}_i} \doteq (\mathbf{H}_{\mathbf{y}_i}\mathbf{H}_{\mathbf{u}_i}^{\perp})(\mathbf{H}_{\mathbf{y}_i}\mathbf{H}_{\mathbf{u}_i}^{\perp})^T \tag{9}$$

If the inputs $\mathbf{u}_i$ excite all modes of the system, then there exists a pair $(\mathbf{A}, \mathbf{C})$ such that (1) is satisfied if and only if there exist two matrices $\mathbf{G}_{\mathbf{v}_1}, \mathbf{G}_{\mathbf{v}_2} \succeq 0$, $\|\mathbf{G}_{\mathbf{v}_i}\|_2 \le \epsilon$, such that

$$rank(\mathbf{G}_{\mathbf{y}_1} - \mathbf{G}_{\mathbf{v}_1}) \le n \tag{10}$$

and

$$\begin{aligned}\lim_{\sigma\to 0} J_{ld}&\left(\tfrac{1}{2}(\mathbf{G}_{\mathbf{y}_1} + \mathbf{G}_{\mathbf{y}_2} - \mathbf{G}_{\mathbf{v}_1} - \mathbf{G}_{\mathbf{v}_2}) + \sigma\mathbf{I},\right. \\ &\left.\mathbf{G}_{\mathbf{y}_1} - \mathbf{G}_{\mathbf{v}_1} + \sigma\mathbf{I}\right) < \infty\end{aligned} \tag{11}$$

The condition above is only necessary for feasibility of Problem 1, since it does not rule out the possibility of trajectories having been generated by two systems having the same $(\mathbf{A}, \mathbf{C})$ pairs but different B. Ruling out this possibility requires strengthening the hypothesis of the lemma above to require using the same input for both experiments and imposing that the input and measurement noise be weakly uncorrelated (in a sense to be precisely defined below).

Lemma 4: Consider two output sequences $(\mathbf{y}_1, \mathbf{y}_2)$, corresponding to the same input $\mathbf{u}$ but different initial conditions $\mathbf{x}_{0,1}$, $\mathbf{x}_{0,2}$. Assume that the input $\mathbf{u}$ and noise $\mathbf{v}$ are uncorrelated, in the sense that $\mathbf{H_v}\mathbf{V} = 0$ where $\mathbf{V}^T$ is a basis for the row space of $\mathbf{H_u}$. Finally, assume that the input $\mathbf{u}$ is persistently exciting and that condition (5) holds. Then, Problem 1 is feasible if and only if

(i) $\mathbf{F}_1 = \mathbf{F}_2$, where $\mathbf{F}_i \doteq \mathbf{H}_{y_i}\mathbf{V}\mathbf{V}^T\mathbf{H}_{y_i}^T,\ i = 1, 2$; and

(ii) there exist two matrices $\mathbf{G}_{\mathbf{v}_1}, \mathbf{G}_{\mathbf{v}_2} \succeq 0$, $\|\mathbf{G}_{\mathbf{v}_i}\|_2 \le \epsilon$, such that (10) and (11) hold.

IV. Behavior (In)Validation as a J_{ld} Minimization Problem

In principle, the results of the previous section allow for finding infeasibility certificates for Problem 1 by solving a sequence of optimization problems and computing:

$$\begin{aligned}\lim\nolimits_{\sigma\to 0} \min_{\mathbf{G}_{\mathbf{v}1}, \mathbf{G}_{\mathbf{v}2}} & J_{ld}(\tfrac{1}{2}(\mathbf{G}_{\mathbf{y}1} + \mathbf{G}_{\mathbf{y}2} - \mathbf{G}_{\mathbf{v}_1} - \mathbf{G}_{\mathbf{v}_2}) + \sigma\mathbf{I}, \\ & \qquad \mathbf{G}_{\mathbf{y}1} - \mathbf{G}_{\mathbf{v}_1} + \sigma\mathbf{I}) \\ \text{subject to: } & \\ & \mathbf{G}_{\mathbf{v}i} \succeq 0,\ \|\mathbf{G}_{\mathbf{v}i}\|_2 \le \epsilon \\ & \text{rank}(\mathbf{G}_{\mathbf{y}1} - \mathbf{G}_{\mathbf{v}1}) < n + 1\end{aligned} \tag{12}$$

From Theorem 2 it follows that Problem 1 is feasible if and only if the optimal value in (12) is finite. However, the problem above is computationally challenging due to the fact that both the objective function and the last constraint are non-convex[3]. To circumvent this difficulty, in the sequel we introduce a relaxation that avoids the rank minimization constraint and it is convex provided that the noise levels are not too large. This relaxation hinges on the following result:

Theorem 3: Given two noisy input/output sequences $(\mathbf{u}_1, \mathbf{y}_1)$ and $(\mathbf{u}_2, \mathbf{y}_2)$, assume, without loss of generality, that $\mathbf{G}_{\mathbf{y}_1}$ and $\mathbf{G}_{\mathbf{y}_2}$ have full rank[4], and that there exists matrices $\mathbf{G}_{\mathbf{v}_i} \succeq 0$, $\|\mathbf{G}_{\mathbf{v}_i}\|_2 \le \epsilon$ such that (10) and (11) hold. Let $\hat{\mathbf{G}}_{\mathbf{y}_1} \doteq \mathbf{G}_{\mathbf{y}_1} - \sigma_{min}\mathbf{u}_{min}\mathbf{u}_{min}^T$ where σ_{min} and $\mathbf{u}_{min}$ denote the minimum singular value of $\mathbf{G}_{\mathbf{y}_1}$ and its corresponding singular vector, respectively. Then,

1.- Problem 1 is solvable only if $\sigma_{min} \le \epsilon$

2.- $\lim_{\sigma\to 0} J_{ld}^*(\sigma) < \infty$ where

$$\begin{aligned}J_{ld}^*(\sigma) \doteq \min\nolimits_{\mathbf{\Phi}} & J_{ld}(\tfrac{1}{2}(\hat{\mathbf{G}}_{\mathbf{y}_1} + \mathbf{G}_{\mathbf{y}2} - \mathbf{\Phi}) + \sigma\mathbf{I}, \\ & \qquad \hat{\mathbf{G}}_{\mathbf{y}_1} + \sigma\mathbf{I}) \\ \text{subject to: } & \\ & \mathbf{\Phi} \succeq 0 \text{ and } \|\mathbf{\Phi}\|_2 \le 3\epsilon\end{aligned} \tag{13}$$

Proof: The first condition is necessary for feasibility of the rank and norm constraints in (12). To prove the second statement, note that if (12) is finite, then there exist some

[2]Note that this distance is independent of the coordinate system chosen, since $J_{ld}(.,.)$ is invariant under similarity transformations.

[3]Since $\mathbf{G}_{\mathbf{v}i}$ are confined to compact sets, the $\lim_{\sigma\to 0} J_{ld}(.,.)$ can be handled by simply solving a sequence of problems with decreasing σ, since the Bolzano-Weierstrass theorem guarantees that the corresponding sequence of solutions have an accumulation point.

[4]This can be always accomplished by choosing the form factor for $\mathbf{H}_{\mathbf{y}i}$.

matrices $\mathbf{G}^*_{\mathbf{v}1}$ and $\mathbf{G}^*_{\mathbf{v}2}$, with $\|\mathbf{G}^*_{\mathbf{v}i}\| \leq \epsilon$ such that

$$\frac{1}{2}(\mathbf{G}_{\mathbf{y}1} - \mathbf{G}^*_{\mathbf{v}_1} + \mathbf{G}_{\mathbf{y}2} - \mathbf{G}^*_{\mathbf{v}_2}) + \sigma\mathbf{I} = \mathbf{U}\begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \sigma\end{bmatrix}\mathbf{U}^T$$

Next, write

$$\begin{aligned} \mathbf{G}_{\mathbf{y}1} - \mathbf{G}^*_{\mathbf{v}_1} + \mathbf{G}_{\mathbf{y}2} - \mathbf{G}^*_{\mathbf{v}_2} &= \\ \hat{\mathbf{G}}_{\mathbf{y}_1} + \mathbf{G}_{\mathbf{y}2} - (\mathbf{G}^*_{\mathbf{v}_2} + \mathbf{G}^*_{\mathbf{v}_1} - \sigma_{min}\mathbf{u}_{min}\mathbf{u}^T_{min}) &= \\ \hat{\mathbf{G}}_{\mathbf{y}_1} + \mathbf{G}_{\mathbf{y}2} - \mathbf{\Phi} & \end{aligned} \tag{14}$$

where we have defined

$$\mathbf{\Phi} \doteq (\mathbf{G}^*_{\mathbf{v}_2} + \mathbf{G}^*_{\mathbf{v}_1} - \sigma_{min}\mathbf{u}_{min}\mathbf{u}^T_{min}).$$

By construction, $\mathbf{G}^*_{\mathbf{v}1} - \sigma_{min}\mathbf{u}_{min}\mathbf{u}^T_{min} \succeq 0$, since $\mathbf{\Psi} \doteq \sigma_{min}\mathbf{u}_{min}\mathbf{u}^T_{min}$ is the smallest matrix that makes $\mathbf{G}_{\mathbf{y}1}$ rank deficient. Let $\mathbf{\Phi}^* \doteq (\mathbf{G}^*_{\mathbf{v}_2} + \mathbf{G}^*_{\mathbf{v}_1} - \sigma_{min}\mathbf{u}_{min}\mathbf{u}^T_{min})$. The proof follows now from the facts that $\|\mathbf{\Phi}^*\|_2 \leq \|\mathbf{G}^*_{\mathbf{v}_2}\|_2 + \|\mathbf{G}^*_{\mathbf{v}_1}\|_2 + \sigma_{min} \leq 3\epsilon$ and that, by construction, $\hat{\mathbf{G}}_{\mathbf{y}_1} + \mathbf{G}_{\mathbf{y}2} - \mathbf{\Phi}$ and $\hat{\mathbf{G}}_{\mathbf{y}_1}$ share the same null space. ■

Remark 1: The advantage of reformulating Problem 1 in terms of the optimization above, instead of (12) stems from the fact that, from Lemma 2, it follows that the problem is convex in the region

$$\frac{\mathbf{G}_{\mathbf{y}2} - \hat{\mathbf{G}}_{\mathbf{y}_1} - \mathbf{\Phi}}{2} \preceq \sqrt{2}(\hat{\mathbf{G}}_{\mathbf{y}_1} + \sigma\mathbf{I})$$

that is, in cases where the trajectories correspond to initial conditions that are not too far apart. Moreover, even outside this region, the problem can be efficiently handled via "convex plus concave" optimization tools [9].

Based on Theorem 3, we propose the following (conceptual) algorithm for behavioral model (in)validation:

Algorithm 1 J_{ld} based behavioral model (in)validation

1: Data: input sequences $\mathbf{u}_1, \mathbf{u}_2$, (noisy) measurements $\mathbf{y}_1, \mathbf{y}_2$. A priori information: noise bound ϵ
2: $\mathbf{G}_{\mathbf{y_1}} \leftarrow \mathbf{H}_{\mathbf{y_1}}\mathbf{H}^{\perp}_{\mathbf{u_1}}(\mathbf{H}_{\mathbf{y_1}}\mathbf{H}^{\perp}_{\mathbf{u_1}})^T$
3: $\mathbf{G}_{\mathbf{y_2}} \leftarrow \mathbf{H}_{\mathbf{y_2}}\mathbf{H}^{\perp}_{\mathbf{u_2}}(\mathbf{H}_{\mathbf{y_2}}\mathbf{H}^{\perp}_{\mathbf{u_2}})^T$
4: Compute $\sigma_{min}(\mathbf{G}_{\mathbf{y_i}}), i = 1, 2.$
5: **if** $\max \sigma_{min}(\mathbf{G}_{\mathbf{y_i}}) > \epsilon$ **then**
6: The given sequences are not behaviors of the same system
7: **else**
8: $\hat{\mathbf{G}}_{\mathbf{y_1}} \leftarrow \mathbf{G}_{\mathbf{y_1}} - \sigma_{min}\mathbf{u}_{min}\mathbf{u}^T_{min}$
9: Compute

$$J_o = \lim_{\sigma \to 0} \min_{\mathbf{\Phi}} J_{ld}\left(\frac{\hat{\mathbf{G}}_{\mathbf{y_1}} + \mathbf{G}_{\mathbf{y_2}} - \mathbf{\Phi}}{2} + \sigma\mathbf{I}, \hat{\mathbf{G}}_{\mathbf{y_1}} + \sigma\mathbf{I}\right)$$

subject to $\mathbf{\Phi} \succeq 0,\ \|\mathbf{\Phi}\|_2 \leq 3\epsilon$

10: **if** $J_o = \infty$ **then**
11: The given sequences are not behaviors of the same system
12: **end if**
13: **end if**

Note that the above algorithm is conceptual, in the sense that it requires computing the limit as $\sigma \to 0$ and establishing that this limit is finite. However, computing J_{ld} for very small values of σ may lead to numerical instabilities. Thus, from a practical standpoint, we will replace the limit above by simply setting σ to a small value σ_o (typically an order of magnitude smaller than the smallest singular value of $\mathbf{H}_{\mathbf{y}_i}$) and using $J_{ld}(\sigma_o)$ as an invalidation certificate. In this setting, the hypothesis that the two trajectories have been generated by the same system is considered to be (in)validated when $J_{ld}(\sigma_o) > J_T$, some suitably chosen threshold. As illustrated in the next section with several examples, this heuristics performs well both in academic examples and non-trivial real ones.

V. Examples

In this section we illustrate the effectiveness of the proposed method with one academic and two practical examples.

A. Example 1

In this example we consider data generated by the impulse response of the systems given below, with random initial conditions.

$$H_1(z) = \frac{z^3}{(z^2-1.3964z+1)(z-0.6557)}$$
$$H_2(z) = \frac{z^4}{(z^2+1.3473z+1)(z^2-1.9107z+1)}$$
$$H_3(z) = \frac{z^4}{(z^2-1.6360z+1)(z^2-1.9107z+1)}$$

The signals y_1 and y_2 were generated by System 1, y_3, y_4 from System 2, and y_5, y_6 from System 3. In each case we constructed a Hankel matrix from each signal with $m = 10$ rows. Thus, the Grammian matrices $\mathbf{G}_\mathbf{y} \in \mathbb{R}^{10\times10}$. Then, we computed, as proposed in the last section, the distances listed in Table I below. These results show that distances between outputs from the same dynamic system are small while the distances between the outputs from different dynamic systems are large.

TABLE I
PAIRWISE DISTANCES BETWEEN SIGNALS FROM DIFFERENT SYSTEMS

Dist	y1	y2	y3	y4	y5	y6
y1	0.0000	0.0104	4.0562	4.2826	2.5451	2.6421
y2	0.0075	0.0000	3.8640	4.0797	2.6439	2.6989
y3	3.1601	3.1619	0.0000	0.0001	1.0905	0.2157
y4	2.9471	2.9525	0.0287	0.0000	1.1796	0.1940
y5	0.7719	0.7669	1.7655	2.0125	0.0000	0.0381
y6	1.1833	1.1829	1.8342	2.0649	0.0323	0.0000

B. Example 2

In this example, we used real data from a computer vision application. The Berkeley Multimodal Human Action Database (MHAD) [5] is a recent, well recorded, database for human action research. It includes 11 actions performed by 12 subjects. Each subject repeated each action five times. The ground-truth data were acquired with an optical motion capture system, Impulse, which captured 3D positions of LED markers on the subjects. With post-processing, 3D 35-joint skeleton trajectories are assigned to each subject.

In our example, we experiment on two subjects with three actions: jumping in place, punching and waving two hands. We concatenated the x, y, z coordinates of 35 joints in each frame into a vector. Thus, the skeleton trajectories of an action formed a 105-dimensional vector sequence. We built a Hankel matrix for each sequence with $m = 4$ (block) rows. Then, we constructed Grammian matrices and computed the distances between the denoised, regularized matrices obtained using the procedure outlined in Section IV. Since the number of frames for each recorded action was different, the lengths of the vector sequences were different. On the other hand, the dimension of the Grammian matrix $\mathbf{H}\mathbf{H}^T$ remains 420×420 for all the sequences. The experimental results are summarized in Table II, where it can be seen that distances between matrices originating in sequences from the same action are small while those corresponding to different actions are large.

TABLE II
PAIRWISE DISTANCES BETWEEN ACTIONS IN MHAD. J: JUMPING IN PLACE; P: PUNCHING; W: WAVING TWO HANDS;

Dist	J1	J2	P1	P2	W1	W2
J1	0.0000	0.0126	0.2463	0.2355	0.2056	0.2107
J2	0.0122	0.0000	0.2486	0.2375	0.1911	0.1974
P1	0.2286	0.2313	0.0000	0.0364	0.1989	0.2016
P2	0.2232	0.2264	0.0413	0.0000	0.2013	0.2016
W1	0.1938	0.1764	0.2433	0.2374	0.0000	0.0077
W2	0.1951	0.1803	0.2428	0.2366	0.0070	0.0000

C. Fault Detection

For the third example, we used the second set of the IMS Bearing Data [7] to show that our metric is able to detect mechanical failures. The experimental setup consists of four bearings installed on a shaft which keeps a constant 2000 RPM rotation with a 6000 lbs radial load. Sensors are placed to record a 1-second vibration signal snapshot every 10 minutes and the system was run until at least one bearing failed.

The data, following common practice, was pre-processed as follows. The original signal was first down-sampled from 20kHz to 1kHz and then the components above 500Hz were filtered out. The resulting pre-processed signal has 984 snapshots, each of which has 4 channels (bearings) and where each of the channels has 1024 data points. Then, we selected Snapshot 1 as an anchor and compared all the other snapshots against it. Only signals of the same channel (bearing) between different snapshots were compared. If there was no failure, the dynamics of each channel should not change and the distance between the anchor and the snapshot should be close to zero. On the other hand, if the distance was large, there was a dynamic change and a failure probably happened. The results of our experiment are shown in Figure 1.

In this dataset, the ground truth is that an outer race failure occurred in Bearing 1 at the end of the test. From the experimental results we can see that indeed Bearing 1 started changing dynamics around Snapshot 700. We also observe that the dynamics of all bearings started to change around Snapshot 900 following the same pattern. A possible explanation for this behavior is that the failure of bearing 1 probably also changed the underlying dynamics of the entire rig.

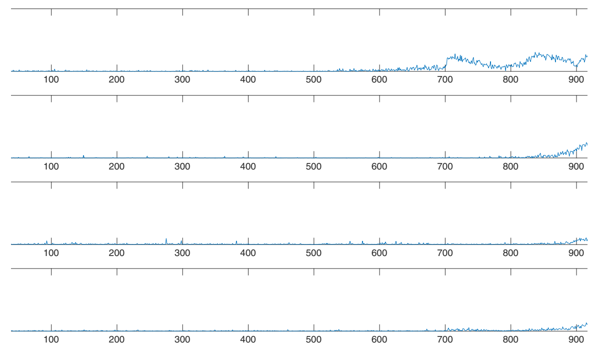

Fig. 1. Distance between each snapshot and the anchor snapshot for each bearing.

VI. CONCLUSIONS

Many problems of practical interest require establishing whether two given input/output trajectories, potentially affected by noise correspond to behaviors of the same (unknown) LTI system. The main result of this paper shows that this problem can be reduced to a Jensen-Bregman divergence minimization form which can be efficiently solved by using recently proposed algorithms. The effectiveness of this approach was illustrated with both, an academic example and two non–trivial problems: activity recognition in video sequences and data–driven fault detection.

REFERENCES

[1] Arun Cherian, Suvrit Sra, Adrish Banerjee, and Nikolaos Papanikolopoulos. Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(9):2161–2174, 2013.

[2] B. Li, O. Camps, and M. Sznaier. Cross-view activity recognition using hankelets. In *IEEE Conf. Comp. Vision and Pattern Recog. (CVPR)*, Providence, RI, June 2012.

[3] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor. *Exact and Approximate Modeling of Linear Systems : A Behavioral Approach*. SIAM, 2006.

[4] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle. On- and off-line identification of linear state space models. *International Journal of Control*, 49(1):219–232, 1989.

[5] Ferda Ofli, Rizwan Chaudhry, Gregorij Kurillo, René Vidal, and Ruzena Bajcsy. Berkeley mhad: A comprehensive multimodal human action database. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 53–60. IEEE, 2013.

[6] F. Paganini. A set based approach for white noise modeling. *IEEE Trans. Autom. Contr.*, 41(10):1453–1465, 1996.

[7] Hai Qiu, Jay Lee, Jing Lin, and Gang Yu. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration*, 289(4):1066–1090, 2006.

[8] M. Sznaier and O. Camps. A rank minimization approach to trajectory (in)validation. In *ACC*, 2011.

[9] Alan L Yuille and Anand Rangarajan. The concave-convex procedure (cccp). *Advances in neural information processing systems*, 2:1033–1040, 2002.