Abstract—This paper considers the problem of determining whether two pairs of (noisy) time sequences \( (u, y) \) 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/output 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

\[
\begin{align*}
\mathbb{R} & \text{ set of real numbers} \\
x, M & \text{ a vector in } \mathbb{R}^n \text{ (matrix in } \mathbb{R}^{n \times m}) \\
M^T & \text{ transpose of matrix } M \\
\sigma_M & \text{ maximum singular value of } M \\
\sigma_{M_{i}} & \text{ }^i\text{th singular value of } M \\
|M| & \text{ determinant of } M \\
M^\perp & \text{ (right) annihilator of } M: MM^\perp = 0 \text{ and } (M^\perp)^TM^\perp = I
\end{align*}
\]
the matrix $M - N$ is positive semidefinite.

$S_n^+$ set of symmetric matrices in $\mathbb{R}^{n \times n}$

cone of positive-semidefinite (definite) matrices in $S_n^+$

$\dim(X)$ dimension of the subspace $X$

$E(X)$ expected value of a stochastic matrix $X$.

$J_{ld}(X, Y)$ Jensen-Bregman Log Det Divergence:

$J_{ld}(X, Y) = \log \left( \frac{X + Y}{2} \right) - \frac{1}{2} \log |XY|, \ X, Y \in S_n^+$

$H_y^{m,n}(T_y^{n})$ Hankel (Toeplitz) matrix associated with a vector sequence $y(.)$:

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

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

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

$\Gamma_{A,C}^T = [C^T (CA)^T \cdots (CA^k)^T \cdots ]^T$

For simplicity, the dimensions of $H$ and $T$ 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:

$x(k) = Ax(k) + Bu(k), \ z(k) = Cx(k), \ y(k) = z(k) + v(k)$

we will denote by $H_G$ and $T_G$ the Hankel and Toeplitz matrices associated with its impulse response sequence $z(k) = CA^kB, \ k = 0, 1, \ldots$

$B. \ Problem \ Statement$

In this paper we consider the problem of determining whether two given (noisy) input/output pairs $(u_1, y_1)$ and $(u_2, y_2)$ are behaviors of the same (unknown) LTI system. That is, whether $z_1 = y_1 - v_1$ and $z_2 = y_2 - v_2$ are solutions of a set of equations of the form (1) for some triple $(A, B, C)$ and suitable initial conditions and noise sequences $v_1, v_2$. Formally, this leads to the following problem:

$Problem \ 1$: Given two input/output pairs $(u_1(k), y_1(k))$, $k = t_1, \ldots, t_{r_1}$ and $(u_2(k), y_2(k))$, $k = t_2, \ldots, t_{r_2}$, and a bound $\eta$ on the covariance of the measurement noise, determine whether there exists a triple $(A, B, C)$, initial conditions $x_{0,1}, x_{0,2}$ and admissible noise sequences $\{v_1(k)\}, \{v_2(k)\}$, $\|E(v_1v_1^T)\|_2 \leq \eta$ such that (1) holds with $z_i(k) = y_i(k) - v_i(k)$

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

$\|G_v\|_2 \leq \epsilon \doteq \frac{1}{n}\eta$  \hspace{1cm} (2)

where $G_v = H_v^T(H_v^T)^+$.  

$C. \ Background \ Results$

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 $(A, B)$ and $(A, C)$ are controllable and observable, respectively, with $A \in \mathbb{R}^{n \times n}$. Then:

$H_v = \Gamma_{A,C}X + T_G H_u + H_v$ \hspace{1cm} (3)

where $X = [x(0) \ x(1) \ \cdots \ x(t)]$. Further, if $span_{row}(X) \cap span_{row}(H_u) = \{0\}$ \hspace{1cm} (4)

then

$rank(H_vH_u^\perp) = rank(X)$ \hspace{1cm} (5)

(or, equivalently, $X^\perp U = 0$ where $U$ is an orthonormal basis of the row space of $H_u^\perp$).

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

$III. \ Main \ Results$

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 $X, Y \in S_n^+$ is finite if and only if these matrices share the same null-space.

$Lemma \ 3$: Let $X, Y \in S_n^+$ and define $X(\sigma) = X + \sigma I$ and $Y(\sigma) = Y + \sigma I$. Assume, without loss of generality, that

$rank(X) = r_x \geq rank(Y) = r_y$. Then $lim_{\sigma \rightarrow 0} J_{ld}(\frac{1}{2}(X+Y), X) < 0 \iff \text{span}_{col}(X^\perp) \subseteq \text{span}_{col}(Y^\perp)$.

1In 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 \((A, C)\) if the \(d_d\) 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 \(H_u\), 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 \((u_i, z_i)\), \(i = 1, 2\), define the matrices
\[
G_{zi} = (H_{zi}H_{u_i}^\perp)(H_{zi}H_{u_i}^\perp)^T
\]
Assume that \(\text{rank}(G_{z1}) = \text{rank}(G_{z2}) = n\) and condition (4) holds. Then,
\begin{enumerate}[(i)]
  \item There exists a pair \((A, C)\) and initial conditions \(x_{1,0}\) and \(x_{2,0}\) such that (1) is satisfied iff
  \[
  \lim_{\sigma \to 0} J_d \left( \frac{1}{2} G_{z1} + G_{z2} + \sigma I, G_{z1} + \sigma I \right) < \infty
  \]
  \[(7)\]
  \item If (7) holds, then
  \[
  \lim_{\sigma \to 0} J_d \left( \frac{1}{2} G_{z1} + G_{z2} + \sigma I, G_{z1} + \sigma I \right) =
  J_d \left( \frac{1}{2} \begin{pmatrix} X_1 H_{u_1}^\perp & (H_{u_1}^\perp)^T X_1^T \end{pmatrix} + \frac{1}{n} \begin{pmatrix} X_2 H_{u_2} & (H_{u_2}^\perp)^T X_2^T \end{pmatrix} \right)
  \]
  \[X_1 H_{u_1} \perp (H_{u_1} \perp)^T \]
  \[(8)\]
\end{enumerate}

where \(X_1, X_2\) 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 \(S_n^m\).

**Theorem 2:** Given two input/output pairs \((u_i, y_i)\), \(i = 1, 2\), define the matrices
\[
G_{y_i} = (H_i H_{u_i}^\perp)(H_i H_{u_i}^\perp)^T
\]
If the inputs \(u_i\) excite all modes of the system, then there exists a pair \((A, C)\) such that (1) is satisfied if and only if there exist two matrices \(G_{v1}, G_{v2}\) \(\geq 0\), \(\|G_{v}\|_2 \leq \epsilon\) such that
\[
\text{rank}(G_{y_i} - G_{v_i}) \leq n
\]
and
\[
\lim_{\sigma \to 0} J_d \left( \frac{1}{2} G_{y1} + G_{y2} - G_{v1} - G_{v2} + \sigma I, G_{y1} - \sigma I \right) < \infty
\]
\[(10)\]

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 \((A, C)\) 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 \((y_1, y_2)\), corresponding to the same input \(u\) but different initial conditions \(x_{0,1}, x_{0,2}\). Assume that the input \(u\) and noise \(v\) are uncorrelated, in the sense that \(H_u V = 0\) where \(V^T\) is a basis for the row space of \(H_u\). Finally, assume that the input \(u\) is persistently exciting and that condition (5) holds. Then, Problem 1 is feasible if and only if
\begin{enumerate}[(i)]
  \item \(F_1 = F_2\), where \(F_i = H_{yi} V V^T H_{yi}^T\), \(i = 1, 2\); and
  \item there exist two matrices \(G_{v1}, G_{v2}\) \(\geq 0\), \(\|G_{v}\|_2 \leq \epsilon\), such that (10) and (11) hold.
\end{enumerate}

**IV. Behavior (In)Validation as a \(J_d\) 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:
\[
\lim_{\sigma \to 0} \min_{G_{v1}, G_{v2}} J_d \left( \frac{1}{2} (G_{y1} + G_{y2} - G_{v1} - G_{v2}) + \sigma I, G_{y1} - G_{v1} + \sigma I \right)
\]
subject to:
\[
G_{v1} \geq 0, \|G_{v}\|_2 \leq \epsilon
\]
\[
\text{rank}(G_{y1} - G_{v1}) < n + 1
\]
\[(12)\]

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 \((u_1, y_1)\) and \((u_2, y_2)\), assume, without loss of generality, that \(G_{y1}\) and \(G_{y2}\) have full rank\(^4\), and that there exists matrices \(G_{v1}\) \(\geq 0\), \(\|G_{v}\|_2 \leq \epsilon\) such that (10) and (11) hold.

Let \(G_{y1} = G_{y1} - \sigma_{\text{min}} u_{\text{min}}\), \(G_{y2} = G_{y2} - \sigma_{\text{min}} u_{\text{min}}\) where \(\sigma_{\text{min}}\) and \(u_{\text{min}}\) denote the minimum singular value of \(G_{y1}\) and its corresponding singular vector, respectively. Then,
\begin{enumerate}[(1.)]
  \item Problem 1 is solvable only if \(\sigma_{\text{min}} \leq \epsilon\)
  \item \(\lim_{\sigma \to 0} J^*_{d}(\sigma) < \infty\) where
    \[
    J^*_{d}(\sigma) = \min_{\sigma} J_d(\frac{1}{2} (G_{y1} + G_{y2} - \Phi) + \sigma I, G_{y1} + \sigma I)
    \]
    subject to:
    \[
    \Phi \geq 0 \text{ and } \|\Phi\|_2 \leq 3\epsilon
    \]
\end{enumerate}
\[(13)\]

**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 \(d_d(\ldots)\) is invariant under similarity transformations.

\(^3\)Since \(G_{v1}\) are confined to compact sets, the \(\lim_{\sigma \to 0} J_d(\ldots)\) can be handled by simply solving a sequence of problems with decreasing \(\sigma\), since the Bolzano-Weierstrass theorem guarantees that the corresponding sequence of solutions has an accumulation point.

\(^4\)This can be always accomplished by choosing the form factor for \(H_y\).
matrices $G^{*}_v_1$ and $G^{*}_v_2$, with $\|G^{*}_v_i\| \leq \epsilon$ such that
\[
\frac{1}{2}(Gy_1 - G^{*}_v_1 + Gy_2 - G^{*}_v_2) + \sigma I = U^T M 0 0 \ldots \]

Next, write
\[
\tilde{G}y_1 - G^{*}_v_1 + Gy_2 - G^{*}_v_2 = \tilde{G}v_1 + Gy_2 - (G^{*}_v_2 + G^{*}_v_1 - \sigma_{min} u_{min} u_{min}^T) = \] (14)

\[
\tilde{G}y_1 + Gy_2 - \Phi
\]

where we have defined
\[
\Phi = (G^{*}_v_2 + G^{*}_v_1 - \sigma_{min} u_{min} u_{min}^T).
\]

By construction, $G^{*}_v_1 - \sigma_{min} u_{min} u_{min}^T \succeq 0$, since $\Psi \succeq \sigma_{min} u_{min} u_{min}^T$ is the smallest matrix that makes $Gy_1$ rank deficient. Let $\Phi^* = (G^{*}_v_2 + G^{*}_v_1 - \sigma_{min} u_{min} u_{min}^T)$. The proof follows now from the facts that $\|\Phi^*\|_2 \leq \|G^{*}_v_2\|_2 + \|G^{*}_v_1\|_2 + \sigma_{min} \leq \epsilon$ and that, by construction, $\tilde{G}y_1 + Gy_2 - \Phi$ and $Gy_2$ share the same null space.

**Remark 1:** The advantage of reformulating Problem 1 in terms of the optimization above, instead of (12) stems from Algorithm 1 based behavioral model (in)validation:

**Algorithm 1** $J_{id}$ based behavioral model (in)validation

1. Data: input sequences $u_1, u_2$, (noisy) measurements $y_1, y_2$. A priori information: noise bound $\epsilon$
2. $Gy_1 \leftarrow H y_1 H + (H y_1 H + )^T$
3. $Gy_2 \leftarrow H y_2 H + (H y_2 H + )^T$
4. Compute $\sigma_{min}(Gy_1), i = 1, 2$.
5. if $\max\sigma_{min}(Gy_1) > \epsilon$ then
6. The given sequences are not behaviors of the same system
7. else
8. $\tilde{G}y_1 \leftarrow G^{*}_v_1 - \sigma_{min} u_{min} u_{min}^T$
9. Compute
\[
J_o = \lim_{\sigma \to 0} \min_{\Phi \succeq 0} J_{id}(\tilde{G}y_1 + \tilde{G}y_2 + \Phi + \sigma I, \tilde{G}y_1 + \sigma I)\]
subject to $\Phi \succeq 0, \|\Phi\|_2 \leq 3\epsilon$
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 $J_{id}$ for very small values of $\sigma$ may lead to numerical instabilities. Thus, from a practical standpoint, we will replace the limit above by simply setting $\sigma$ to a small value $\sigma_o$ (typically an order of magnitude smaller than the smallest singular value of $H y_1$) and using $J_{id}(\sigma_o)$ 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 $J_{id}(\sigma_o) > J_f$, 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.

\[
H_1(z) = \frac{z^2 + 1.3964 + 1}{z^2 - 0.6557}
\]

\[
H_2(z) = \frac{z^2 + 1.3473 + 1}{z^2 - 1.9107 + 1}
\]

\[
H_3(z) = \frac{z^2 + 1.6360 + 1}{z^2 - 1.9107 + 1}
\]

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 $G_y \in R^{10 \times 10}$. 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>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise distances between signals from different systems</td>
</tr>
<tr>
<td>Dist</td>
</tr>
<tr>
<td>$y_1$</td>
</tr>
<tr>
<td>$y_2$</td>
</tr>
<tr>
<td>$y_3$</td>
</tr>
<tr>
<td>$y_4$</td>
</tr>
<tr>
<td>$y_5$</td>
</tr>
<tr>
<td>$y_6$</td>
</tr>
</tbody>
</table>

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 \( \mathbf{HH}^T \) remains \( 420 \times 420 \) 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;

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

#### 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 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.

#### 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


