A Rank Minimization Approach to Trajectory (In)Validation

Mario Sznaier   Octavia Camps

Electrical and Comp. Engineering Department,
Northeastern University,
Boston, MA 02115.

Abstract— This paper addresses the problem of establishing whether two vector time sequences could have been generated by the same (unknown) linear time invariant system, possibly affected by bounded model uncertainty and measurement noise. This problem arises in multiple contexts, including, among others, behavioral systems model (in)validation, determining the minimum number of models needed to cover the set of operating points of a piecewise–linear plant and in several computer vision and image processing problems. The main result of the paper shows that this problem can be reduced to a rank-minimization form and efficiently solved by using recently proposed convex relaxations. These results are illustrated with both a theoretical example and two non–trivial computer vision problems: activity recognition in video sequences and textured image classification.

I. INTRODUCTION AND MOTIVATION

Many problems of practical interest involve determining whether two given output trajectories could have been generated by the same underlying linear time invariant (LTI) system. For instance, when performing identification of piecewise affine plants, it is of interest to determine the minimum number of LTI subsystems required to describe the system. In turn, this entails determining whether data generated by different experiments, possibly at different operating points, can be explained by a single LTI model. A similar problem arises in computer vision, in the context of classifying activities in video sequences. By assuming that similar activities correspond to the output of the same LTI system excited by a suitable input (see for instance [1], [11]), the problem reduces to establishing whether given sequences are indeed the output, possibly corrupted by noise, of the same system. Finally, in the context of behavioral approach to control [14], this problem can be considered as the counterpart to classical model (in)validation, where one seeks now to (in)validate the hypothesis that two given sequences are behaviors of the same dynamical system.

In principle, given the input/output sequences $(\mathbf{u}_1, \mathbf{y}_1)$, $(\mathbf{u}_2, \mathbf{y}_2)$ and a-priori information about the class of models $\mathcal{S}$ and measurement noise $\mathcal{N}$, the problem above can be solved by computing the consistency sets $\mathcal{T}_i$, $i = 1, 2$ corresponding to each pair. e.g. the set of all models and associated initial conditions consistent with the a-priori information that could have generated the pairs, and checking whether their intersection is empty. However, characterizing these sets

is, in general, highly non-trivial, since they are not jointly convex in the models and initial conditions[1].

Alternatively, working in a behavioral context, it is possible to find first the most powerful unfalsified model [10] corresponding to one of the sequences and then check whether the second sequence is indeed a behavior of this model. A problem here is that, due to the presence of measurement noise, this approach may fail to correctly label (noisy) behaviors from the same system.

To avoid these difficulties, in this paper, building on earlier work from realization theory [9] and subspace identification methods [12], we first propose a rank condition to invalidate the hypothesis that two given trajectories originate from the same LTI system. Briefly, we show that, under mild conditions, two input/output trajectories originate from the same system if and only if the matrix obtained by concatenating suitable projections of the Hankel matrices corresponding to each trajectory has the same rank as each of the individual matrices. In the case of noisy measurements and model uncertainty, a necessary and sufficient invalidation certificate can be obtained by minimizing the rank of this matrix with respect to the noise and model uncertainty. Characterizing the latter by means of a semi–definite constraint, coupled with recently introduced convex rank relaxations, allows to reduce the problem to a semidefinite convex optimization that can be efficiently solved using commonly available software. These results are illustrated with both, a theoretical example and two non-trivial computer vision applications: activity recognition from video sequences and texture classification.

II. PRELIMINARIES

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

A. Notation

$\mathbf{x}, \mathbf{M}$	a vector in $\mathbb{R}^n$ (matrix in $\mathbb{R}^{n \times m}$)
$\mathbf{A}^T$	transpose of matrix $\mathbf{A}$.
$\text{trace}\{\mathbf{M}\}$	trace of the matrix $\mathbf{M}$
$\mathbf{I}$	identity matrix
$\mathbf{M} \succeq \mathbf{N}$	the matrix $\mathbf{M} - \mathbf{N}$ is positive semidefinite.

---

[1]Indeed, most of the model (in)validation literature [4], [6], [13], [15], [18], [20] assumes zero initial conditions in order to get tractable problems.

$\bar{\sigma}(\mathbf{M})$	maximum singular value of $\mathbf{M}$.
$\mathcal{BX}(\gamma)$	γ-ball in a normed space $\mathcal{X}$: $\mathcal{BX}(\gamma) = \{x \in \mathcal{X} : \|x\|_{\mathcal{X}} \le \gamma\}$.
$\mathcal{H}_\infty$	space of functions with bounded analytic continuation inside the unit disk, equipped with the norm: $\|X\|_{\mathcal{H}_\infty} \doteq \text{ess } \sup_{\|z\|<1} \bar{\sigma}(X(z))$.
$\mathbf{H}_y^{m,n}$	Hankel matrix associated with a vector sequence $\mathbf{y}(.)$:

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

$\mathbf{T}_y^n$	Toeplitz matrix associated with a vector sequence $\mathbf{y}(.)$:

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

In the sequel, when the dimensions are clear from the context, we will simply denote $\mathbf{H}_y^{m,n}$ and $\mathbf{T}_y^n$ as $\mathbf{H}_y$ and $\mathbf{T}_y$, respectively.

In this paper we consider Linear Time Invariant (LTI) systems with the usual state space representation:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k, \\ \mathbf{z}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k \end{aligned} \qquad (1)$$

In the sequel, for notational simplicity, given a system G of the form (1) we will denote by

$$\mathbf{\Gamma}_G \doteq \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^i \\ \vdots \end{bmatrix} \quad \mathbf{T}_G = \begin{bmatrix} \mathbf{D} & 0 & \cdots & 0 & \cdots \\ \mathbf{CB} & \mathbf{D} & \cdots & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \cdots \\ \mathbf{CA}^i\mathbf{B} & \mathbf{CA}^{i-1}\mathbf{B} & \cdots & \mathbf{D} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \ddots \end{bmatrix} \qquad (2)$$

the matrices that map its initial condition $\mathbf{x}_o$ and input sequence u to its output $\mathbf{z}$, respectively.
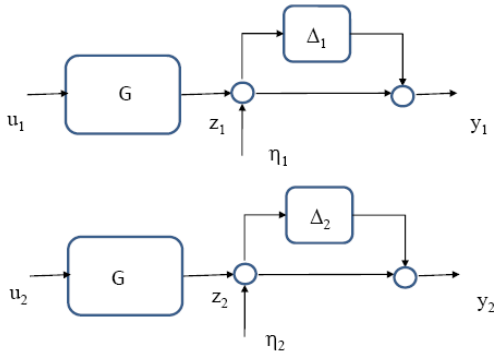
B. Problem Statement

Fig. 1. Model (In)Validation Set–up

In this paper we consider the problem of (in)validating the hypothesis that two given input/ouput pairs $(\mathbf{u}_1, \mathbf{y}_1)$ and $(\mathbf{u}_2, \mathbf{y}_2)$ are generated by the same underlying system,

possibly affected by uncertainty and noise. The setup is illustrated in Fig. 1: here $\mathbf{u}(.)$ and $\mathbf{y}(.)$ represent an input to the (unknown) LTI system G and the corresponding output, affected by uncertainty Δ and measurement noise η. For instance, in the applications of interest to this paper, Δ models the variability intrinsic to two different realizations of the same texture or the same activity, performed at different times. In the sequel we will assume that the only information known about the noise η and model uncertainty are set-membership descriptions, e.g. $\eta \in \mathcal{N}$, $\Delta \in \mathcal{D} = \mathcal{BH}_\infty(\gamma)$, for some given $\mathcal{N}$ convex, and $\gamma > 0$. In this context, the problem under consideration can be precisely stated as:

Problem 1: Given two input/output pairs $(\mathbf{u}_1(t), \mathbf{y}_1(t))$, $t = t_1, \ldots, t_{T_1}$ and $(\mathbf{u}_2(t), \mathbf{y}_2(t))$, $t = t_2, \ldots, t_{T_2}$, a bound n_G on the McMillan degree of G, and the admissible noise and uncertainty sets $\mathcal{N}$, $\mathcal{D}$, determine whether there exists at least one pair $(\eta_1, \eta_2) \in \mathcal{N}$, one pair $(\Delta_1, \Delta_2) \in \mathcal{D}$, a LTI system G with McMillan degree at most n_G, and suitable initial conditions $\mathbf{x}_1$, $\mathbf{x}_2$, such that:

$$\begin{aligned} \mathbf{z}_1(t) &= \mathbf{\Gamma}_G\mathbf{x}_1 + \mathbf{T}_G\mathbf{u}_1(t) \\ \mathbf{z}_2(t) &= \mathbf{\Gamma}_G\mathbf{x}_2 + \mathbf{T}_G\mathbf{u}_2(t) \\ \mathbf{y}_1(t) &= (1 + \Delta_1) * (\mathbf{z}_1(t) + \eta_1(t)), \; t = t_1, \ldots, t_{T_1} \\ \mathbf{y}_2(t) &= (1 + \Delta_2) * (\mathbf{z}_2(t) + \eta_2(t)), \; t = t_2, \ldots, t_{T_2} \end{aligned} \qquad (3)$$

where $*$ denotes convolution.

C. Background Results

Next, we introduce some background results that will be used to solve Problem 1. To this effect, let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ be a minimal realization of G, with $\mathbf{A} \in R^{n \times n}$. Then, it can be shown that the input/output sequences satisfy [12]:

$$\mathbf{H}_y = \mathbf{\Gamma}_G\mathbf{X} + \mathbf{T}_G\mathbf{H}_u \qquad (4)$$

where $\mathbf{H}_u$ and $\mathbf{H}_y$ denote the Hankel matrices associated with the input and output sequences, respectively, $\mathbf{X} \doteq \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \ldots & \mathbf{x}(t) \end{bmatrix}$ is a matrix containing the consecutive state vectors, and where $\mathbf{\Gamma}_G$ and $\mathbf{T}_G$ are defined in (2).

Lemma 1: Let $\mathbf{H}_u^\perp$ denote the right annihilator of $\mathbf{H}_u$ and assume that

$$span_{row}(\mathbf{X}) \cap span_{row}(\mathbf{H}_u) = \{0\} \qquad (5)$$

Then $\text{rank}(\mathbf{H}_y\mathbf{H_u}^\perp) = \text{rank}(\mathbf{X})$.

Proof: Follows from the proof of Theorem 2 in [12].

Corollary 1: If $\text{rank}(\mathbf{X}) = n$ (that is, the input u excites all modes of the system), then $\text{rank}(\mathbf{H_y}\mathbf{H_u}^\perp) = n$

III. MAIN RESULTS

Here we present our main result, a rank-minimization based invalidation certificate. To this effect, we examine first the case where the exact system output is available.

A. Behavior (In)Validation as a Rank Minimization Problem

Lemma 2: Consider two input/output pairs $(\mathbf{u}_1, \mathbf{z}_1)$ and $(\mathbf{u}_2, \mathbf{z}_2)$ and assume that the inputs $(\mathbf{u}_1, \mathbf{u}_2)$ are sufficiently rich so that all modes of the system are excited. Then, there

exist an LTI system G with McMillan degree at most n_G and two initial conditions $\mathbf{x}_1, \mathbf{x}_2 \in R^n$ and such that:

$$\mathbf{z}_1 = \mathbf{\Gamma}_G \mathbf{x}_1 + \mathbf{T}_G \mathbf{u}_1, \ \mathbf{z}_2 = \mathbf{\Gamma}_G \mathbf{x}_2 + \mathbf{T}_G \mathbf{u}_2 \qquad (6)$$

only if the following conditions hold:

(i) $n \doteq \mathrm{rank}(\mathbf{H}_{\mathbf{z}_1}\mathbf{H}_{\mathbf{u}_1}{}^\perp) = \mathrm{rank}(\mathbf{H}_{\mathbf{z}_2}\mathbf{H}_{\mathbf{u}_2}{}^\perp) \le n_G$
(ii) $n \doteq \mathrm{rank}(\mathbf{H}) \doteq \mathrm{rank}([\mathbf{H}_{\mathbf{z}_1}\mathbf{H}_{\mathbf{u}_1}{}^\perp \quad \mathbf{H}_{\mathbf{z}_2}\mathbf{H}_{\mathbf{u}_2}{}^\perp]) \le n_G$

Further, in this case all LTI systems of degree $\le n_G$ that could have generated the experimental data share the same pair $(\mathbf{A}, \mathbf{C})$ up to a similarity transformation.

Proof: If (6) holds for some G with McMillan degree $n \le n_G$, then, from Lemma 1 it follows that:

$$\mathrm{rank}(\mathbf{H}_{\mathbf{z}_1}\mathbf{H}_{\mathbf{u}_1}{}^\perp) = \mathrm{rank}(\mathbf{H}_{\mathbf{z}_2}\mathbf{H}_{\mathbf{u}_2}{}^\perp) = n \le n_G \qquad (7)$$

Moreover, from (4) we have:

$$\mathbf{H} = \begin{bmatrix} \mathbf{\Gamma}_G & \mathbf{\Gamma}_G \end{bmatrix} \begin{bmatrix} \mathbf{X}_1\mathbf{H}_{u_1}^\perp & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2\mathbf{H}_{u_2}^\perp \end{bmatrix} \qquad (8)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are the state trajectories corresponding to the inputs $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. If (5) holds then the matrix $\begin{bmatrix} \mathbf{X}_1\mathbf{H}_{u_1}^\perp & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2\mathbf{H}_{u_2}^\perp \end{bmatrix}$ has full row rank. It follows that:

$$\mathrm{rank}(\mathbf{H}) = \mathrm{rank}\left(\begin{bmatrix} \mathbf{\Gamma}_G & \mathbf{\Gamma}_G \end{bmatrix} \right) = n \qquad (9)$$

To complete the proof, note that if (i) holds, then there exist two LTI systems G_1, G_2 with McMillan degree at most n, and initial conditions $\mathbf{x}_1, \mathbf{x}_2$ such that [12]:

$$\mathbf{z}_1 = \mathbf{\Gamma}_{G_1}\mathbf{x}_1 + \mathbf{T}_{G_1}\mathbf{u}_1, \ \mathbf{z}_2 = \mathbf{\Gamma}_{G_2}\mathbf{x}_2 + \mathbf{T}_{G_2}\mathbf{u}_2 \qquad (10)$$

Let $(\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1)$ and $(\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2)$ be (minimal) state space realizations of G_1 and G_2, respectively. If (ii) and (5) hold, then

$$\mathrm{rank}\left(\begin{bmatrix} \mathbf{\Gamma}_{G_1} & \mathbf{\Gamma}_{G_2} \end{bmatrix} \right) = n \Rightarrow \mathbf{\Gamma}_{G_1} = \mathbf{\Gamma}_{G_2}\mathbf{M} \qquad (11)$$

for some full rank matrix $\mathbf{M}$. Hence

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{C}_2\mathbf{M} \\ \mathbf{C}_1\mathbf{A}_1 &= \mathbf{C}_2\mathbf{A}_2\mathbf{M} \\ &\vdots \\ \mathbf{C}_1\mathbf{A}_1^{t-1} &= \mathbf{C}_2\mathbf{A}_2^{t-1}\mathbf{M} \end{aligned} \qquad (12)$$

Let $\hat{\mathbf{A}}_2 \doteq \mathbf{M}^{-1}\mathbf{A}_2\mathbf{M}$. It can be easily shown that (12) implies that $\mathbf{C}_1\mathbf{A}_1^{k-1} = \mathbf{C}_1\mathbf{A}_1^{k-2}\hat{\mathbf{A}}_2$, $k = 2, \ldots, t$. Thus

$$\begin{bmatrix} \mathbf{C}_1\mathbf{A}_1 \\ \mathbf{C}_1\mathbf{A}_1^2 \\ \vdots \\ \mathbf{C}_1\mathbf{A}_1^n \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_1\mathbf{A}_1 \\ \vdots \\ \mathbf{C}_1\mathbf{A}_1^{n-1} \end{bmatrix} \mathbf{A}_1 = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_1\mathbf{A}_1 \\ \vdots \\ \mathbf{C}_1\mathbf{A}_1^{n-1} \end{bmatrix} \hat{\mathbf{A}}_2 \qquad (13)$$

which (from observability of $(\mathbf{A}_1, \mathbf{C}_1)$) implies that $\mathbf{A}_1 = \hat{\mathbf{A}}_2$. It follows then that the pairs $(\mathbf{A}_1, \mathbf{C}_1)$ and $(\mathbf{A}_2, \mathbf{C}_2)$ are related by the similarity transformation $\mathbf{M}$. ∎

Note that the result above does not rule out the possibility of the data having been generated by two different systems with the same pair $(\mathbf{A}, \mathbf{C})$ but different $\mathbf{B}$ and $\mathbf{D}$. A simple example is the impulse response of two systems, one with

$\mathbf{D}_1 = 0$ and initial condition $\mathbf{x}_o$, and the second with initial condition 0, $\mathbf{B}_2 = \mathbf{A}\mathbf{x}_o$ and $\mathbf{D}_2 = \mathbf{C}\mathbf{x}_o$. Thus, in the sequel, to prevent this situation, we will strengthen condition (5) to

$$\mathbf{X} \cdot \mathbf{V}^T = 0 \qquad (14)$$

where $\mathbf{V}$ is a basis for the row space of $\mathbf{H}_{\mathbf{u}}{}^2$, and consider the scenario where the same input is used in both experiments. With these assumptions, Lemma 2 can be strengthened to:

Theorem 1: Consider an input signal $\mathbf{u} \ne 0$, and two output sequences $(\mathbf{z}_1, \mathbf{z}_2)$, corresponding to state trajectories $\mathbf{X}_1$ and $\mathbf{X}_2$. Assume that $\mathbf{u}$ is sufficiently exciting and that $\mathbf{X}_i$ satisfy condition (14). Then, there exists an LTI system G with McMillan degree at most n_G and such that:

$$\mathbf{z}_1 = \mathbf{\Gamma}_G \mathbf{x}_1 + \mathbf{T}_G \mathbf{u}, \ \mathbf{z}_2 = \mathbf{\Gamma}_G \mathbf{x}_2 + \mathbf{T}_G \mathbf{u} \qquad (15)$$

if and only if the following conditions hold:

(i) $n \doteq \mathrm{rank}(\mathbf{H}_{\mathbf{z}_1}\mathbf{H}_{\mathbf{u}}{}^\perp) = \mathrm{rank}(\mathbf{H}_{\mathbf{z}_2}\mathbf{H}_{\mathbf{u}}{}^\perp) \le n_G$
(ii) $n \doteq \mathrm{rank}(\mathbf{H}) \doteq \mathrm{rank}([\mathbf{H}_{\mathbf{z}_1}\mathbf{H}_{\mathbf{u}}{}^\perp \quad \mathbf{H}_{\mathbf{z}_2}\mathbf{H}_{\mathbf{u}}{}^\perp]) \le n_G$
(iii) $(\mathbf{H}_{z_1} - \mathbf{H}_{z_2})\mathbf{V}^T = 0$.

Proof: Necessity of (i) and (ii) is immediate from Lemma 2. Necessity of (iii) follows from the fact that if (14) and (15) hold then

$$(\mathbf{H}_{z_1} - \mathbf{H}_{z_2})\mathbf{V}^T = \mathbf{\Gamma}_G(\mathbf{X}_1 - \mathbf{X}_2)\mathbf{V}^T = 0$$

To show sufficiency, let G_1 and G_2 be two LTI systems of McMillan degree n such that (10) holds, with corresponding realizations $(\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1, \mathbf{D}_1)$ and $(\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2, \mathbf{D}_2)$. From Lemma 2 it follows that these realizations can be selected so that $\mathbf{A}_1 = \mathbf{A}_2$ and $\mathbf{C}_1 = \mathbf{C}_2$. To complete the proof, we need to show that in this coordinate system $\mathbf{B}_1 = \mathbf{B}_2$ and $\mathbf{D}_1 = \mathbf{D}_2$. Note that if (14) holds then $(\mathbf{X}_1 - \mathbf{X}_2)\mathbf{H}_{\mathbf{u}}{}^T = 0$. Similarly, if (iii) holds then $(\mathbf{H}_{z_1} - \mathbf{H}_{z_2})\mathbf{H}_{\mathbf{u}}{}^T = 0$. Hence

$$\begin{aligned} 0 &= \mathbf{\Gamma}(\mathbf{X}_1 - \mathbf{X}_2)\mathbf{H}_u^T + (\mathbf{T}_{G_1} - \mathbf{T}_{G_2})\mathbf{H}_u\mathbf{H}_u^T \\ &= (\mathbf{T}_{G_1} - \mathbf{T}_{G_2})\mathbf{H}_u\mathbf{H}_u^T \end{aligned} \qquad (16)$$

Assume that $\mathbf{u}(0) \ne 0$. Then, since the first row of $(\mathbf{T}_{G_1} - \mathbf{T}_{G_2})$ has the form $[\mathbf{D}_1 - \mathbf{D}_2, 0, \ldots, 0]$ it follows that $\mathbf{D}_1 = \mathbf{D}_2$. The same argument successively applied to the rows of $(\mathbf{T}_{G_1} - \mathbf{T}_{G_2})$ shows that (16) implies that $\mathbf{C}\mathbf{A}^i(\mathbf{B}_1 - \mathbf{B}_2) = 0$, $i = 1, 2, \ldots, n$. The fact that $\mathbf{B}_1 = \mathbf{B}_2$ follows now from observability of $(\mathbf{A}, \mathbf{C})$, as long as $m \ge n$. Finally, note that if $\mathbf{u}(0) = 0$, a similar reasoning can be applied starting with the first non-zero element of the input sequence. ∎

B. (In)Validation Certificates for noisy sequences

Next, we recast Problem 1 into a rank minimization. Consider first a simpler case without model uncertainty, e.g:

Problem 2: Given an input $\mathbf{u}(t)$ and measurements $\mathbf{y}_i(t)$, $i = 1, 2$ of the corresponding outputs $\mathbf{z}_i(t)$ corrupted by noise $\eta_i(t) \in \mathcal{N}$, determine whether there exists a pair $(\eta_1, \eta_2) \in \mathcal{N}$, a LTI system G with McMillan degree at most n_G and initial conditions $\mathbf{x}_1, \mathbf{x}_2$ such that:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{\Gamma}_G \mathbf{x}_1 + \mathbf{T}_G \mathbf{u} + \eta_1 \\ \mathbf{y}_2 &= \mathbf{\Gamma}_G \mathbf{x}_2 + \mathbf{T}_G \mathbf{u} + \eta_2 \end{aligned} \qquad (17)$$

---

[2]Intuitively, this condition imposes that the state at a time k is uncorrelated with values of the input $\mathbf{u}(t)$, $t \ge k$.

where $\mathbf{y}_i^T = \left[\mathbf{y}_i(t), \ldots, \mathbf{y}_i(t+T)\right]^T$, $\mathbf{u}_i^T = \left[\mathbf{u}(t), \ldots, \mathbf{u}(t+T)\right]^T$, and $\eta_i^T = \left[\eta_i(t), \ldots, \eta_i(t+T)\right]^T$. From Theorem 1, it follows that Problem 2 is feasible iff there exist noise sequences $\eta_1, \eta_2 \in \mathcal{N}$ such that

$$n \doteq \mathrm{rank}\left((\mathbf{H}_{y_1} - \mathbf{H}_{\eta_1})H_u^\perp\right) =$$
$$\mathrm{rank}\left((\mathbf{H}_{y_2} - \mathbf{H}_{\eta_2})H_u^\perp\right) \leq n_G$$
$$n \doteq \mathrm{rank}\left(\left[(\mathbf{H}_{y_1} - \mathbf{H}_{\eta_1})\mathbf{H}_u^\perp \quad (\mathbf{H}_{y_2} - \mathbf{H}_{\eta_2})\mathbf{H}_u^\perp\right]\right) \leq n_G$$
$$(\mathbf{H}_{y_2} - \mathbf{H}_{\eta_2} - \mathbf{H}_{y_2} + \mathbf{H}_{\eta_2})\mathbf{V}^T = 0$$

Note that, in practical situations n_G is usually unknown and must be also estimated from the available data. Thus, in the sequel we will make the additional assumption that all the Hankel singular values of the system are larger than the noise level (an observability assumption that guarantees that the correct model order can be estimated from the data). Under this additional assumption certifying the hypothesis that the two given signals were generated by the same underlying system can be accomplished proceeding as follows:

Algorithm 1: *Conceptual Behavior Model (In)Validation*

Data: input sequence u, (noisy) measurements y_1, y_2.
A priori information: noise description $\eta_i \in \mathcal{N}$
1. Solve the following rank–minimization problems:
   $r_1^{min} = \min_{\eta_1} \mathrm{rank}((\mathbf{H}_{y_1} - \mathbf{H}_{\eta_1})\mathbf{H}_u^\perp)$ s.t. $\eta_1 \in \mathcal{N}$.
   $r_2^{min} = \min_{\eta_2} \mathrm{rank}((\mathbf{H}_{y_2} - \mathbf{H}_{\eta_2})\mathbf{H}_u^\perp)$ s.t. $\eta_2 \in \mathcal{N}$.
   $r_{12}^{min} = \min_{\eta_1} \mathrm{rank}([\mathbf{H}_{y_{1n}}\mathbf{H}_u^\perp \quad \mathbf{H}_{y_{2n}}\mathbf{H}_u^\perp])$
   s.t $\eta_1, \eta_2 \in \mathcal{N}$, $\mathbf{H}_{y_{1n}} = \mathbf{H}_{y_1} - \mathbf{H}_{\eta_1}$,
   $\mathbf{H}_{y_{2n}} = \mathbf{H}_{y_2} - \mathbf{H}_{\eta_2}$ and $(\mathbf{H}_{y_{1n}} - \mathbf{H}_{y_{2n}})\mathbf{V}^T = 0$
   where $\mathbf{V}$ is a basis for the row space of $\mathbf{H}_u$
2. The given trajectories were generated by the same LTI system iff $r_1^{min} = r_2^{min} = r_{12}^{min}$

Algorithm 1 is conceptual, in the sense that it entails solving affine matrix rank–minimization problems, subject to linear constraints, a problem that is generically NP–hard. To obtain tractable problems, motivated by the heuristics in [3], [7][3], we will relax problems of the form:

$$\min_{\mathbf{x}} \mathrm{rank}(\mathbf{H}(\mathbf{x})) \text{ subject to } \mathbf{L}(\mathbf{x}) = 0$$

where $\mathbf{H}(.)$ and $\mathbf{L}(.)$ are affine functions of $\mathbf{x}$ to the following sequence of convex problems:

$$\min_{\mathbf{x},\mathbf{Y},\mathbf{Z}} \quad \mathrm{trace}(\mathbf{W}_y^{(k)}\mathbf{Y}^{(k)}) + \mathrm{trace}(\mathbf{W}_z^{(k)}\mathbf{Z}^{(k)})$$
$$\text{subject to} \quad \begin{bmatrix} \mathbf{Y}^{(k)} & \mathbf{H}^{(k)} \\ (\mathbf{H}^{(k)})^T & \mathbf{Z}^{(k)} \end{bmatrix} \succeq 0$$
$$\mathbf{L}(\mathbf{x}) = 0$$
(18)

where $\mathbf{Y}^{(k)} \succeq 0$ and $\mathbf{Z}^{(k)} \succeq 0$ are auxiliary variables and $\mathbf{W}_z^{(k)}, \mathbf{W}_y^{(k)}$ are weighting matrices. At the kth iteration, these weights are selected according to the formula $\mathbf{W}_y^{(k)} = (\mathbf{Y}^{(k-1)} + \lambda_{k-1}\mathbf{I})^{-1}$, $\mathbf{W}_z^{(k1)} = (\mathbf{Z}^{(k-1)} + \lambda_{k-1}\mathbf{I})^{-1}$, where

---

[3]Although there are very recent faster algorithms for rank minimization (e.g. [2]), they currently cannot handle semidefinite constraints.

$\mathbf{Y}^{(k-1)}, \mathbf{Z}^{(k-1)}$ denote the optimal solution to (18) at the $(k-1)^{th}$ iteration and where the regularization parameter λ_{k-1} is set to the $(h - n_q + 1)^{th}$ largest singular value of $\mathbf{H}^{(k)}$ [4]. Since the matrices $\mathbf{H}$ and $\mathbf{L}$ are affine in optimization variables $\mathbf{x}$ and $\mathbf{Y}^{(k)}$ and $\mathbf{Z}^{(k)}$ are symmetric positive definite auxiliary variables, (18) is a convex semidefinite program that can be solved using widely available software.

Using this relaxation in Algorithm 1 leads to the following convex optimization based (in)validation algorithm:

Algorithm 2: *Connvex Behavior Model (In)Validation*

Data: input sequence u, (noisy) measurements y_1, y_2.
A priori information: noise description $\eta_i \in \mathcal{N}$
1. Use the relaxation (18) to (approximately) solve the following rank–minimization problems:
   $r_1^* = \min_{\eta_1} \mathrm{rank}((\mathbf{H}_{y_1} - \mathbf{H}_{\eta_1})\mathbf{H}_u^\perp)$ s.t. $\eta_1 \in \mathcal{N}$.
   $r_2^* = \min_{\eta_2} \mathrm{rank}((\mathbf{H}_{y_2} - \mathbf{H}_{\eta_2})\mathbf{H}_u^\perp)$ s.t. $\eta_2 \in \mathcal{N}$.
   $r_{12}^* = \min_{\eta_1} \mathrm{rank}([\mathbf{H}_{y_{1n}}\mathbf{H}_u^\perp \quad \mathbf{H}_{y_{2n}}\mathbf{H}_u^\perp])$
   s.t $\eta_1, \eta_2 \in \mathcal{N}$, $\mathbf{H}_{y_{1n}} = \mathbf{H}_{y_1} - \mathbf{H}_{\eta_1}$,
   $\mathbf{H}_{y_{2n}} = \mathbf{H}_{y_2} - \mathbf{H}_{\eta_2}$ and $(\mathbf{H}_{y_{1n}} - \mathbf{H}_{y_{2n}})\mathbf{V}^T = 0$
   where $\mathbf{V}$ is a basis for the row space of $\mathbf{H}_u$
2. If $r_1^* = r_2^* = r_{12}^*$ then the given trajectories are admissible behaviors of the same system G, when starting from suitable initial conditions.

Note that $r_i^* \geq r_i^{min}$ and $r_{12}^* \geq r_{12}^{min}$ due to the use of the convex relaxation in Algorithm 2. Hence failure of the condition $r_1^* = r_2^* = r_{12}^* = n_G$ does not necessarily invalidates the hypothesis that the output trajectories are (noisy) behaviors of the same system.

C. Handling model uncertainty

Consider again Problem 1, where, in addition to noise, the trajectories are affected by bounded model uncertainty $\Delta \in \mathcal{BH}_\infty(\gamma)$ where γ is given as part of a-priori information. In this scenario, the internal signal $\mathbf{z}$ is given by:

$$\mathbf{z}(t) = \zeta(t) - \eta(t), \ \eta \in \mathcal{N} \tag{19}$$

where the signal ζ satisfies:

$$\mathbf{y} = (1 + \Delta) * \zeta, \text{ for some } \Delta \in \mathcal{D} \tag{20}$$

where $*$ denotes convolution. From Theorem 2.3.6 in [5], a necessary and sufficient condition for feasibility of (20) is[5]:

$$\gamma^2 \mathbf{T}_\zeta^T \mathbf{T}_\zeta - (\mathbf{T}_y - \mathbf{T}_\zeta)^T(\mathbf{T}_y - \mathbf{T}_\zeta) \succeq 0 \tag{21}$$
$$\Updownarrow$$
$$\begin{bmatrix} \mathbf{T}_y^T\mathbf{T}_y - \mathbf{T}_y^T(\mathbf{T}_z + \mathbf{T}_\eta) - (\mathbf{T}_z + \mathbf{T}_\eta)^T\mathbf{T}_y & (\mathbf{T}_z + \mathbf{T}_\eta)^T \\ \mathbf{T}_z + \mathbf{T}_\eta & \frac{1}{1-\gamma^2} \end{bmatrix} \succeq 0 \tag{22}$$

where the last line follows from a Schur complement argument. Since (22) is a Linear Matrix inequality in $\mathbf{z}, \eta$, it

---

[4]The matrices $\mathbf{W}_y^{(0)}, \mathbf{W}_z^{(0)}$ are set to the identity. Hence the first iteration solves the nuclear norm heuristic. Then each iteration aims to reduce the rank further through the weighting scheme. In our experiments, convergence is typically achieved within the first 10 iterations.

[5]This follows from application of the Theorem to the signal ζ and $\mathbf{y} - \zeta$.

follows that the only modification to Algorithm 2 required to handle model uncertainty, is to incorporate this additional (convex) constraint to the rank minimization problems.

IV. EXAMPLES

In this section we illustrate our results with a theoretical example and two non-trivial computer vision applications: texture image classification and activity recognition.

A. Theoretical Example

In this simple example we consider the zero initial condition step responses $z_1(t)$ and $z_2(t)$ of the systems $G_1(\lambda) = \frac{\lambda+1}{(\lambda^2+w_1^2)(\lambda^2+w_2^2)}$ and $G_2(\lambda) = \frac{\lambda}{(\lambda^2+w_1^2)(\lambda^2+w_3^2)}$ with $w_1 = exp(j0.5)$, $w_2 = exp(j0.4)$, and $w_3 = exp(j0.7)$. The "measured" data consists of four sequences, obtained from portions of these trajectories corrupted by noise as follows:

$$
\begin{aligned}
y_{11}(t) &= z_1(t) + \eta_1, \ t \in [1, 25] \\
y_{12}(t) &= z_1(t) + \eta_1, \ t \in [41, 65] \\
y_{21}(t) &= z_2(t) + \eta_2, \ t \in [1, 25] \\
y_{22}(t) &= z_2(t) + \eta_2, \ t \in [41, 65]
\end{aligned}
$$

In order to apply Algorithm 2, a noise characterization is required. Let $\mathbf{H}_\eta$ denote the Hankel matrix of the noise sequence. Note that the (i, j) element of $\mathbf{H}_\eta^{\mathbf{T}}\mathbf{H}_\eta$ is given by:

$$
\mathbf{H}_\eta^{\mathbf{T}}\mathbf{H}_\eta(i, j) = \sum_{r=1}^{n} \eta(i+r)\eta(j+r)
$$

Thus, under mild ergodicity assumptions $\frac{1}{n}\mathbf{H}_\eta^{\mathbf{T}}\mathbf{H}_\eta$ is an estimate of the noise covariance matrix. Thus, constraints on $\|\mathbf{H}_\eta\|$ are (approximately) equivalent to constraints on the magnitude of the noise covariance. Hence, in the sequel we will consider noise sets of the form:

$$
\mathcal{N} = \{\eta \colon \frac{1}{n}\bar{\sigma}(\mathbf{H}_\eta) \leq \epsilon\}
$$

where ϵ is given as part of the a-priori information. In particular, in this example ϵ was set to $0.1 * \frac{1}{n}\mathbf{H}_y$ (e.g. 10% noise level). The results of applying Algorithm 2 (with no model uncertainty) are shown in Table I. As shown there, the algorithm correctly identifies y_{11} and y_{12} as being generated by the same system. On the other hand, as expected, while rank($\mathbf{H}_{y11}$) =rank($\mathbf{H}_{y22}$) = 4, rank($[\mathbf{H}_{y11} \ \mathbf{H}_{y22}]$) > 4.

TABLE I

HANKEL RANK FOR THE THEORETICAL EXAMPLE

Sequence Pair	rank($\mathbf{H}_1$)	rank($\mathbf{H}_2$)	rank($[\mathbf{H}_1 \ \mathbf{H}_2]$)
(y_{11}, y_{12})	4	4	4
(y_{11}, y_{22})	4	4	5

B. Texture Classification

The purpose of this example is to illustrate the application of the proposed framework to the problem of texture classification. This problem has been the subject of intense research in the computer vision and image processing communities, with application ranging from medical diagnosis to object recognition and image database retrieval. Most
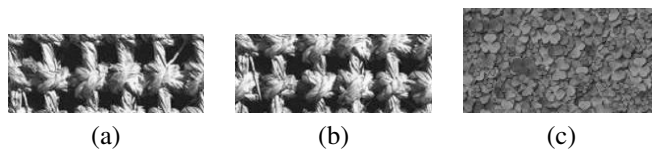

(a)   (b)   (c)

Fig. 2.   Three sample patches used for recognition.

texture recognition schemes are stochastic in nature, relying on representations in terms of statistics of the responses to a collection of filters [8]. Here, we pursue an alternative approach, based upon recasting the problem into a robust behavior (in)validation form of the form shown in Figure 1. Motivated by the work in [19], we will postulate that all images corresponding to realizations of a given texture $\mathcal{T}$ are realizations of a second order stationary random process. Further, for simplicity, we will assume uniformity in the vertical direction. In this context, the kth column of the image can be assumed to be the output, at index k, of an unknown one-dimensional system G to a random initial condition x_o and the problem of determining whether or not two given images are samples of the same texture reduces to Problem 1[6].

The approach outlined above was tested on the textured images shown in Figure 2. As illustrated in Table II, the algorithm correctly identifies (a) and (b) as slices of the same texture.

TABLE II

TEXTURE CLASSIFICATION RESULTS

Image Pair	rank($\mathbf{H}_1$)	rank($\mathbf{H}_2$)	rank($[\mathbf{H}_1 \ \mathbf{H}_2]$)
(a, b)	10	10	10
(a, c)	10	10	13

C. Activity Classification

This example explores the application of the proposed framework to the problem of recognizing human activities from video sequences. For this purpose, we used 2 video sequences, a walking and a running person, from the KTH database [17]. Sample frames from these sequences are shown in Figure 3. In order to reduce the dimensionality of the data, the videos were pre-processed by first cropping each frame to the size of a bounding box around the actor of the activity, resizing all the frames so they are all the size of the largest bounding box (104×42 pixels), and then projecting the frames into a three dimensional space using principal component analysis (PCA) of the pair of videos being compared.

Applying Algorithm 2 to the PCA time traces of the PCA coefficients shown in Figure 4, assuming 10% noise and 10% model uncertainty, led to the results shown in Table III. As shown there, the algorithm correctly identifies

---

[6]The uniformity assumption can be removed by considering 2D Roesser models [16] and extending the theory presented here to this case. However, this is beyond the scope of the present paper.
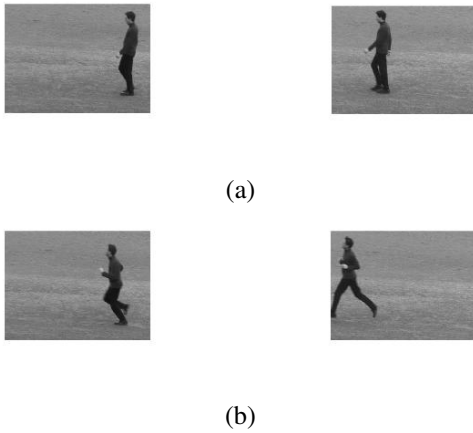
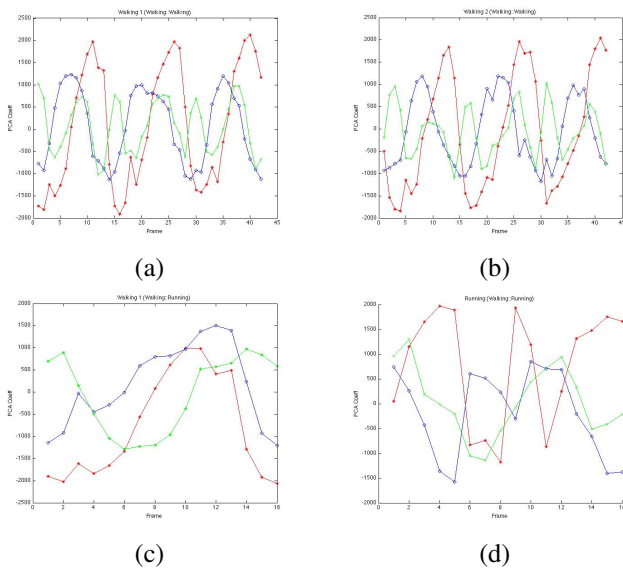Fig. 3. Sample frames from KTH activity video database. (a) Walking. (b) Running.

Fig. 4. PCA coefficients for 3 subsequences: (a), (b), (c) walking and (d) running.

the subsequences (a),(b),(c) as being generated by the same underlying activity (walking).

V. CONCLUSIONS

Many problems of practical interest require establishing whether two given input/output trajectories, potentially affected by noise and uncertainty, correspond to behaviors of the same (unknown) LTI system. The main result of this paper shows that this problem can be reduced to a rank-minimization form and efficiently solved by using recently proposed convex relaxations. The effectiveness of this approach was illustrated with both, a theoretical example and two non–trivial computer vision problems: activity recognition in video sequences and textured image classification.

REFERENCES

[1] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, December 2001.

[2] J. F. Cai, E. J. Candes, and Z Shen. A singular value thresholding algorithm for matrix completion. preprint.

[3] E. J. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.

[4] J. Chen. Frequency-domain tests for validation of linear fractional uncertain models. *IEEE Transactions on Automatic Control*, 42(6):748–760, 1997.

[5] J. Chen and G. Gu. *Control Oriented System Identification, An $\mathcal{H}_\infty$ Approach*. John Wiley, New York, 2000.

[6] J. Chen and S. Wang. Validation of linear fractional uncertain models: Solutions via matrix inequalities. *IEEE Transactions on Automatic Control*, 41(6):844–849, 1996.

[7] M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of American Control Conf. 2003*, volume 3, pages 2156–2162. AACC, 2003.

[8] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

[9] B.L. Ho and R.E. Kalman. Effective construction of linear, state-variable models from input/output functions. *Regelungstechnik*, 14:545–548, 1966.

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

[11] M. C. Mazzaro, M. Sznaier, and O. Camps. A model (in)validation approach to gait classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(11):1820–1825, 2005.

[12] 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.

[13] M. P. Newlin and R. S. Smith. A generalization of the structured singular value and its application to model validation. *IEEE Transactions on Automatic Control*, 43(7):901–907, 1998.

[14] J.W. Polderman and J.C. Willems. *Introduction to Mathematical Systems Theory: A Behavioral Approach*. Springer, New York, 1998.

[15] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal. A time domain approach to model validation. *IEEE Transactions on Automatic Control*, 39(5):951–959, 1994.

[16] R. P. Roesser. A discrete state-space model for linear image processing. *IEEE Trans. on Automatic Control*, 20(1):1–10, Feb. 1975.

[17] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: A local svm approach. In *In Proc. ICPR*, pages 32 – 36, 2004.

[18] R. S. Smith and J. C. Doyle. Model validation: A connection between robust control and identification. *IEEE Trans. Autom. Contr.*, 37(7):942–952, 1992.

[19] M. Sznaier, O. I. Camps, and C. Mazzaro. Finite horizon model reduction of a class of neutrally stable systems with applications to texture synthesis and recognition. In *IEEE Conference on Decision and Control*, Paradise Island, Bahamas, Dec. 2004.

[20] D. Xu, Z. Ren, G. Gu, and J. Chen. LFT uncertain model validation with time and frequency–domain measurements. *IEEE Transactions on Automatic Control*, 44(7):1435–1441, 1999.

TABLE III

HUMAN ACTIVITY CLASSIFICATION RESULTS. SEQUENCES (A),(B),(C) CORRESPOND TO WALKING AND SEQUENCE (D) TO RUNNING

Activity Pair	rank($\mathbf{H}_1$)	rank($\mathbf{H}_2$)	rank($[\mathbf{H}_1\ \mathbf{H}_2]$)
(a,b)	4	4	4
(a,c)	4	4	4
(a,d)	4	8	8