AFRL-AFOSR-VA-TR-2020-0080

Surviving the Data Deluge: A Combined Dynamical Systems/Machine Learning Approach

Mario Sznaier
NORTHEASTERN UNIVERSITY

12/19/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
25-06-2020	Final Performance	01 Sep 2015 to 31 Aug 2019

4. TITLE AND SUBTITLE
Surviving the Data Deluge: A Combined Dynamical Systems/Machine Learning Approach

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0392

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Mario Sznaier

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NORTHEASTERN UNIVERSITY
360 HUNTINGTON AVE
BOSTON, MA 02115 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2020-0080

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This research sought to develop a comprehensive, computationally tractable framework for synthesizing information driven systems capable of both autonomously operating and supporting safety--critical human operations in rapidly changing ``data deluged' scenarios. Its conceptual backbone was a rigorous integration of systems theory, machine learning and optimization elements that emphasized robustness, computational simplicity and improved situational awareness. The research advanced the state of the art in systems theory by developing a tractable framework for robust identification/model (invalidation) of a broad class of dynamical systems that incorporates ideas from machine learning and semi-algebraic optimization to handle outliers, missing data and substantial noise levels.

15. SUBJECT TERMS
Dynamical System Identification, Machine Learning and Control

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			LEVE, FREDERICK
Unclassified	Unclassified	Unclassified	UU		19b. TELEPHONE NUMBER *(Include area code)* 703-696-7309

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Surviving the Data Deluge: A Combined Dynamical Systems/Machine Learning Approach.

Mario Sznaier
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Phone (617) 373–5364, Fax (617) 373–8970
email: msznaier@coe.neu.edu

Octavia I. Camps
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Phone (617) 373–4663, Fax (617) 373–8970
email: camps@coe.neu.edu

December 1, 2019

Final Report, Grant FA9550-15-1-0392
Dynamics and Control Program

Air Force Office of Scientific Research
Program Officer: Dr. Frederick A. Leve, AFOSR/RT.

Project Summary

Scope: This research sought to develop a comprehensive, computationally tractable framework for synthesizing information driven systems capable of both autonomously operating and supporting safety–critical human operations in rapidly changing "data deluged" scenarios. Its conceptual backbone was a rigorous integration of systems theory, machine learning and optimization elements that emphasized robustness, computational simplicity and improved situational awareness.

Relevance to the USAF mission: Flexible, provably correct autonomy is a key enabler for maintaining the superiority and expanding the capabilities of the USAF in the next two decades. Autonomous systems endowed with analysis and decision capabilities can collect data, assess intention, and if necessary, take action, while at the same time substantially reducing the required manpower and cost, vis-a-vis existing unmanned vehicles. Arguably, a major road-block to realizing this vision stems from the curse of dimensionality. Simply put, existing inferencing techniques are ill-equipped to deal with the extremely large volume of data that needs to be analyzed in real time. This is precisely the challenge addressed by this research: development of a computationally tractable framework that supports provably correct inferencing and decision making in "data deluged" scenarios. The long term vision was to lay the foundations for designing systems endowed with provably correct flexible autonomy, capable of making decisions in-situ, with minimal human intervention, and of supporting human decisions by providing enhanced situational awareness.

Contributions to Basic Science: This research effort sought a rapprochement between systems theory and machine learning, by using the metaphor of dynamic models as parsimonious information encapsulators. It advanced the state of the art in systems theory by developing a tractable framework for robust identification/model (invalidation) of a broad class of dynamical systems that incorporates ideas from machine learning and semi-algebraic optimization to handle outliers, missing data and substantial noise levels. On the other hand, embedding machine learning problems in the conceptual world of dynamical systems lead to scalable, computationally tractable algorithms, that are provably convergent. For instance, in this context, the problems of simultaneous manifold embedding and classification of dynamic data, and semi-supervised learning were recast as a sequence of convex optimization problems that exploited the inherent sparse structure of the problem to substantially reduce the computational burden.

Benefits to the General Public: In addition to directly supporting the USAF mission, autonomous platforms endowed with activity analysis capabilities can assist law enforcement, allow elderly people to continue living independently, and help first responders and emergency workers in preventing hazards from developing into full blown catastrophic situations.

1 Motivation

This research sought to develop a comprehensive, computationally tractable framework for synthesizing information driven systems capable of both autonomously operating and supporting safety–critical human operations in rapidly changing "data deluged" scenarios. Its conceptual backbone was a rigorous integration of systems theory, machine learning and optimization elements that emphasizes robustness, computational simplicity and improved situational awareness. It included both theoretical developments at the confluence of systems theory and machine learning, and an investigation of implementation issues.

1.1 Needs, Transformative Impact and Relevance to the USAF Mission

Flexible, provably correct autonomy is a key enabler for maintaining the superiority and expanding the capabilities of the USAF in the next two decades: Autonomous systems endowed with analysis and decision capabilities can collect data, assess intention, and if necessary, take action, while at the same time substantially reducing the required manpower and cost, vis-a-vis existing unmanned vehicles. Arguably, a major road-block to realizing this vision stems from the curse of dimensionality, illustrated in Figure 1. Simply put, existing inferencing techniques are ill-equipped for analyzing the "data deluge" generated by the sensors, within the constraints imposed by the need for robust, real time operation in dynamic, partially stochastic scenarios. This research effort sought to address this issue by exploiting recent advances in robust dynamical systems, machine learning, semi-algebraic geometry and optimization. The long term vision was to lay the foundations for designing systems endowed with provably correct flexible autonomy, capable of making decisions in-situ, without human intervention, while passing on to the next decision level only mission–relevant situational abstractions.

(a) (b) (c)

Figure 1: Sample scenarios where autonomous systems endowed with analysis capabilities can prevent incidents from developing into tragedies. (a) finding objects left behind; (b) detecting an attempt to breach a "sterile area" by entering through an exit; and (c) detecting an anomalous flight pattern. In all cases decisions must be taken based on events discernible only in a small fraction of a very large data record.

The main idea driving this research was to combat the "curse of dimensionality" by exploiting the twin "blessings" of self-similiarity and concentration of measure. While these ideas had already been recently exploited in machine learning (for instance in the context of dimensionality reduction and variable selection), we sought to extend them to a larger class of problems, inferencing from dynamic data, by embedding the problem in the expanded dynamical systems identification framework outlined in Fig. 2. Briefly, in this approach, self-similarity refers to the observation that actionable data exhibits high degree of spatio-temporal correlation, and therefore can be treated as the output of an underlying switched nonlinear dynamical system, with jumps indicating the occurrence of events. Similarly, concentration of measure, refers to the fact that typically only relatively few parameters (far fewer than the dimensionality of the data), are needed

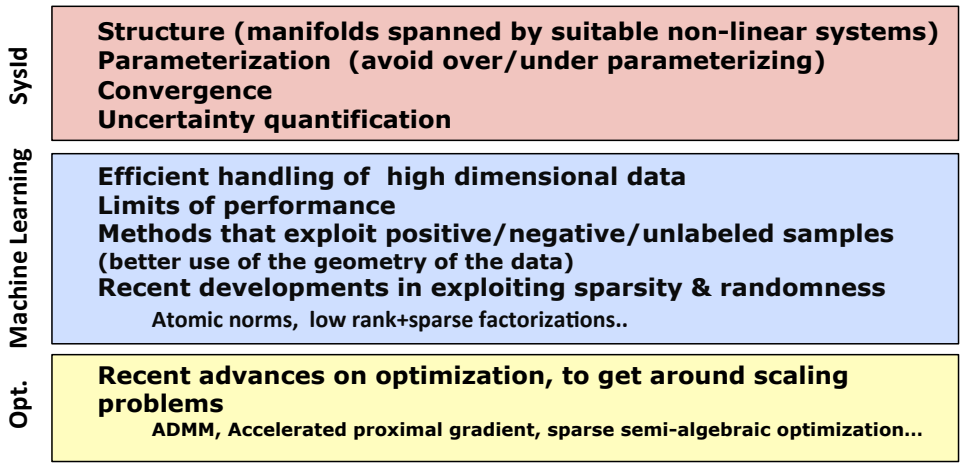


Figure 2: Combating the "curse of dimensionality" using the twin "blessings" of self-similarity and concentration of measure: a combined systems-identification/machine learning approach, enabled by recent advances in sparse optimization.

to completely characterize these models. Combining these observations allowed for recasting the dynamic inferencing problem into a dynamic sparsification form, which in turn could be reduced to a tractable convex semidefinite optimization problem, via recents results connecting semi-algebraic geometry to the classical theory of moments. As illustrated in the sequel, this approach led to compact representations of both the data and underlying dynamics, allowing for computational efficient solutions to problems involving complex dynamics and large sized data vectors.

2 Description of the Basic Research Performed and Summary of the Results

In this section we give a brief summary of the basic research performed under this grant and our findings, grouped into four research themes (i) identification, (ii) model validation, (iii) fault detection, and (iv) control. A more complete description can be found in the papers listed in the publications section, which can be downloaded from the Robust Systems Lab website: http://robustystems.coe.neu.edu.

2.1 Identification.

In principle, embedding information extraction problems in the conceptual world of systems identification made available a rich, extremely powerful resource base, leading to computationally tractable, robust solutions. However, successful application of this idea hinged upon the development of computationally tractable solutions to the following problems, open at the time that the project was started:

2.1.1 Identification of parsimonious dynamical models [1–4]. As indicated in the introduction, the goal of this research was to exploit the ability of sparsification based techniques to provide computationally tractable solutions to hard problems. One such problem is identifying parsimonious models from a combination of noisy measurements and a-priori information, related to the physics of the specific domain under consideration. The main motivation for seeking low order models is the fact that, as we have shown during

this research, computational complexity of both model identification and controller synthesis is directly determined by the model order (since this order determines the tree width of the underlying correlative sparsity graph). Indeed, exploiting the underlying sparsity, it is possible to design identification algorithms that scale linearly with the number of data points, thus opening up the possibility of on-line controllers that make optimal use of all the available information.

(i) Atomic norm approach to identification of LTI systems in the presence of missing data [1,3,4]

In view of the discussion above, our initial goal was to solve the following problem:

Problem 1. *Given N (noisy) samples of the time response y_t of an unknown plant G to a known input u_t and some* a priori *information about the underlying process, (i) determine whether there exists a model compatible with these priors that explains the observed experimental data, and, (ii) if so, find the coefficients of the lowest order model with this property.*

The main approach that we used to solve the problem above was to recast it into an atomic-norm constrained minimization. In this context, a given object (in this case the unknown system to be identified) is expressed in terms of the elements of a given dictionary $\mathcal{A} = \{a\}$ containing the "atoms" a, and assigned an "atomic norm" defined by $\|\mathbf{x}\|_{\mathcal{A}} = \inf\left\{\sum_{a\in\mathcal{A}}|c_a| \ : \mathbf{x} = \sum_{a\in\mathcal{A}} c_a a\right\}$, allowing for reducing the identification problem to a constrained optimization of the form:

$$\min_x f(x) \text{ subject to } \|x\|_{\mathcal{A}} \leq \tau \tag{1}$$

where $f(x)$ is a smooth convex function that measures fidelity to the observed data and τ is used to promote sparsity. Note that (1) can be considered a constrained version of a regularized problem of the form:

$$\min_x f(x) + \lambda\|x\|_{\mathcal{A}} \tag{2}$$

For instance, in the case where the set of atoms consists of basis vectors and their symmetric images with respect to the origin, $\pm e_i \in \mathbb{R}^n$, the corresponding atomic norm is simply the ℓ_1 norm and the problem above reduces to the well known form $\min_x f(x) + \lambda\|x\|_1$. Similarly, a set of of atoms consisting of all unit norm rank-1 matrices leads to problems of the form $\min_x f(x) + \lambda\|x\|_*$. In general scenarios, the set $\mathcal{A}$ can be tailored to include all the a-priori available information, for instance by constraining it to contain only elements compatible with the physics of the problem under consideration. For instance, if a bound ρ on the time constant of the relevant process is known, a suitable set of atoms is given by $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, where:

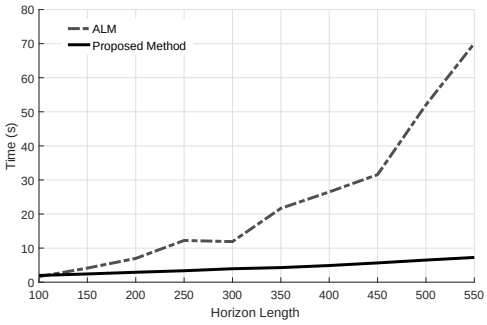


Figure 3: Atomic norm versus ADMM computational complexity scaling.

$$\begin{aligned}
\mathcal{A}_1 &= \left\{\Psi_p(z) = \pm\frac{(1-|p|^2)}{2}\left(\frac{1}{z-p} + \frac{1}{z-p^*}\right) \ : \ |p| \leq \rho\right\} \\
\mathcal{A}_2 &= \left\{\Psi_p(z) = \pm\frac{(1-|p|^2)}{2}\left(\frac{-j}{z-p} + \frac{j}{z-p^*}\right) \ : \ |p| \leq \rho\right\} \\
\mathcal{A}_3 &= \{\Psi_p(z) = \pm 1\} \\
\mathcal{A}_4 &= \left\{\Psi_p(z) = \pm\frac{(1-|p|^2)}{z-p} \ : \ p \in [-\rho, \rho]\right\}
\end{aligned} \tag{3}$$

p^* denotes the complex conjugate of p, and where the normalization factor $1-|p|^2$ guarantees that each Ψ_p has norm less than or equal to 1. A potential difficulty here is that the set $\mathcal{A}$ above is infinite dimensional, since it contains all points inside $\mathbb{D}_\rho$, the origin centered disk of radius ρ. In principle, this could be handled by simply discretizing this disk, but this approach can lead to poor performance if the system to be identified is lightly damped. To avoid this discretization, we have developed the randomized Frank-Wolfe type algorithm shown below:

Algorithm 1 Randomized algorithm to minimize a convex function f over the τ-scaled atomic norm ball

$\mathbf{x_0} \leftarrow \tau\Upsilon^N\{a_0(z)\}$ for arbitrary $a_0(z) \in \mathcal{A}$ ▷ Init.
for k = 0,1,2,3,..., k_{max} **do**
  Select N_k poles uniformly distributed over $\mathbb{D}_\rho$,
  denote the set of these poles S_k
  $\mathbf{a_k} \leftarrow \Upsilon^N\{\text{argmin}_{a(z)\in\mathcal{A}\{S_k\}}\langle\nabla f(x_k), \Upsilon^N\{a(z)\}\rangle\}$
  $\alpha_k \leftarrow \text{argmin}_{\alpha\in[0,1]} f(\mathbf{x_k} + \alpha[\tau\mathbf{a_k} - \mathbf{x_k}])$
  $\mathbf{x_{k+1}} \leftarrow \mathbf{x_k} + \alpha_k[\tau\mathbf{a_k} - \mathbf{x_k}]$
end for

As we have shown in [1,4,5], this algorithm requires computing only inner products and thus its computational complexity scales linearly with the number of data points (see Fig. 3) as opposed to the cubic scale when using ADMMs. Hence, it can comfortably handle very large data sets. Further, this algorithm has a convergence rate of $\mathcal{O}(\frac{1}{t})$ and can be easily modified to handle missing data, non-uniform sampling and multiple runs. Finally, we have used this approach to design a convolutional neural network that performs system identification and time series prediction [6].

(ii) Extension to Wiener models [2]. Wiener systems, the interconnection of a linear time invariant system and a static nonlinearity shown in Fig 4, are interesting both in their own right, since they are pervasive in many diverse domain applications, and as a tractable approximation to general nonlinear control problems. Thus, during the past decade a large research effort has been devoted to the problem of identifying models of these systems from experimental data, leading to a number of techniques. Existing approaches can be roughly divided into statistical and set membership, or control oriented. The latter are attractive for control applications, since the results of the identification process (a nominal plant and bounds on the identification error) can be directly used by robust control methods. Unfortunately, as shown by the PIs, set membership identification of Wiener systems is NP hard, both in the number of experiments and in the number of outputs of the LTI sub-system, *even in cases where the non-linearity is known*. Thus, handling moderately large problems requires the use of relaxations, for instance based on polynomial optimization. However, while these relaxations work well for medium size problems, the entailed computational complexity is non trivial.

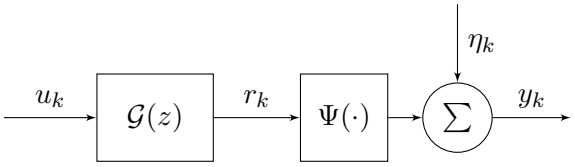


Figure 4: General Wiener identification setup

To circumvent this difficulty, in this research we developed a computationally efficient approach to the problem of set membership identification of the linear portion of a Wiener system from noisy output data, assuming that the non-linearity is known. Such problems arise frequently in control applications, where the physics of the problem determine the nonlinearity. In addition, similar problems arise in the context of manifold embedding of dynamic data, [7] where the embedding function is determined using

machine learning techniques such as Locally Linear Embeddings (LLE) and the challenge is to identify the manifold dynamics. Briefly, the main idea was to use a combination of set inversions and properties of interval matrices to efficiently determine a set of intervals guaranteed to contain, at all times, the output of the (unknown) LTI system. This step was then followed by a combination of binary optimization and atomic norm minimization to identify both the interval containing the actual signal as well as the lowest order LTI system compatible with this choice and any available priors. Our main result (see [2] for details) showed that exploiting the properties of interval matrices to prune the set of candidate intervals, resulted in a substantial complexity reduction (typically 98% to 99% of the possible combinations are discarded), allowing for efficiently solving problems whose size challenges existing techniques.

The starting point for this approach was the following result on interval matrices:

Theorem 1. *The following are sufficient conditions for the regularity of an interval matrix:*

1. *Let A_c be nonsingular and $\rho(|A_c^{-1}|\Delta) < 1$ hold. Then $[A_c - \Delta, A_c + \Delta]$ is regular.*
2. *Let the matrix $A_c^T A_c - \|\Delta^T \Delta\| I$ be positive definite for some consistent matrix norm $\|\cdot\|$. Then $[A_c - \Delta, A_c + \Delta]$ is regular.*

This result allows for substantially pruning the set of candidate trajectories proceeding as follows. Specifically, consider a single input single output LTI system of McMillan degree p, driven by an input sequence $\mathbf{u} = [u_0, u_1, \cdots, u_{M-1}]^T$. Assume that $M \geq 3p+2$ input/output pairs have been collected in the vectors $\mathbf{u} = [u_0, u_1, \cdots, u_{M-1}]^T$ and $\mathbf{y} = [y_0, y_1, \cdots, u_{M-1}]^T$, respectively, and that by performing a one to many set inversion of the nonlinearity, a candidate consistent sequence of intervals for the intermediate signal $\mathbf{r}$, i.e. $\mathcal{V}_{1:M}$, was identified. From well known results in subspace identification methods, it follows that, for the correct interval, the matrix $\mathbf{G} \doteq [\mathbf{H} \;\; \mathbf{U}]$ is rank deficient, where

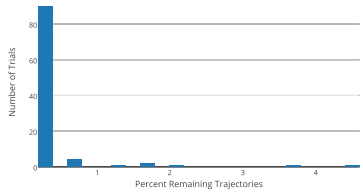
$$\mathbf{H} \doteq \begin{pmatrix} [\underline{r}, \overline{r}]_0 & [\underline{r}, \overline{r}]_1 & \dots & [\underline{r}, \overline{r}]_p \\ [\underline{r}, \overline{r}]_1 & [\underline{r}, \overline{r}]_2 & \dots & [\underline{r}, \overline{r}]_{p+1} \\ \vdots & \ddots & \ddots & \vdots \\ [\underline{r}, \overline{r}]_{2p+1} & [\underline{r}, \overline{r}]_{2p+2} & \dots & [\underline{r}, \overline{r}]_{3p+1} \end{pmatrix}, \; \mathbf{U} \doteq \begin{pmatrix} u_0 & u_1 & u_2 & \dots & u_p \\ u_1 & u_2 & u_3 & \dots & u_{p+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ u_{2p+1} & u_{2p+2} & u_{2p+3} & \dots & u_{3p+1} \end{pmatrix} \tag{4}$$

Thus, all interval matrices $\mathbf{G}$ that are regular do not correspond to valid trajectories and can be eliminated, leading to a set of *unfalsified trajectories*. Finally, the correct trajectory can be obtained by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{c},\mathbf{s}} \quad & \|\mathbf{c}\|_{\ell_1} \\ \text{subject to} \quad & \mathbf{T_u D_a c} = \mathbf{r} \\ & \underline{\mathbf{K}}\mathbf{s} \leq \mathbf{r} \leq \overline{\mathbf{K}}\mathbf{s} \\ & \{\mathbf{s}\}_i \in \{0, 1\} \\ & \sum_i \{\mathbf{s}\}_i = 1 \end{aligned} \tag{5}$$

Here $\mathbf{D_a}$ denotes a dictionary of atomic impulse responses, $\underline{\mathbf{K}}$ and $\overline{\mathbf{K}}$ are matrices whose columns contain the lower and upper bound of the unfalsified intervals and the variable $\mathbf{s} \in \mathbb{R}^J$ selects exactly one interval trajectory to be used as a bound for the estimated output of the LTI block, i.e. $\mathbf{r}$. The resulting problem can be efficiently solved using commercially available solvers such as Gurobi. The effectiveness

of the proposed algorithm is illustrated in Fig. 5 for the case of a third order system cascaded with the (non-invertible) non-linearity $\Psi(r) = r^2$. As shown there, the proposed interval approach falsifies more than 99% of the trajectories, allowing to identify the system in a few seconds (see [2] for more details).



	Minimum	Average	Maximum
Estimation % Error	0.19	1.17	6.20
Estimation Time (s)	0.48	1.16	9.37

Figure 5: Wiener SysId using interval algebra. Left: number of pruned trajectories. Right: identification error

(iii) Extension to LPV and non-linear systems [8, 9]. A common description of dynamic phenomena is a nonlinear state-space model where the set of inputs and states define its operating conditions. When this model is linearized, the resulting linear models are dependent upon the operating points about which the linearization occurred. In a linear parameter varying (LPV) modeling approach, the dependence of the linear model on its operating point is projected into a lower dimensional space called the scheduling space. This projection is guided by physical intuition, or an analysis of the manner in which the inputs and states affect the output. This task is not always apparent and bad assumptions, such as choosing scheduling variables based on data acquisition convenience, may cause loss of fidelity in capturing the observed behavior. In this portion of the research we extended the concept of parsimonious models to LPV and to bilinear systems (by reducing the later case to the former). The main result of this portion of the research showed that both the scheduling parameters and a parsimonious model can be identified by solving a rank minimization problem. Specifically, we considered an ARX structure with time varying parameters:

$$A(t,q)y(t) = B(t,q)u(t) + e(t) \tag{6}$$

where $A(t,q)$ and $B(t,q)$ are time-varying polynomials in delay operator q^{-1}:

$$\begin{aligned} A(t,q) &= 1 + a_1(t)q^{-1} + a_2(t)q^{-2} + \ldots + a_{na}(t)q^{-na} \\ B(t,q) &= b_1(t)q^{-nk} + b_2(t)q^{-nk-1} + \ldots + b_{nb}(t)q^{-nb-nk+1} \end{aligned} \tag{7}$$

nk denotes input-to-output lag which can be zero. The model's parameter vector is:

$$\Theta(t) = [a_1(t), a_2(t), \ldots, a_{na}(t), b_1(t), \ldots, b_{nb}(t)]^T \tag{8}$$

The model structure can also be written as:

$$y(t) = \Theta(t)^T \Phi(t) + e(t) \tag{9}$$

where $\Phi(t)$ is the vector of model's regressors composed of lagged input-output variables. The length of $\Phi(t)$ is $n = na + nb$. The models parameters $\Theta(t)$ are assumed to evolve according to an affine auto-regressive process driven by the "inputs" $u(t)$ and $y(t)$:

$$F(q)(\Theta(t) - \bar{\Theta}) = G_1(q)u(t) + G_2(q)y(t) \tag{10}$$

where $F(q)$, $G_1(q)$, $G_2(q)$ are constant-coefficient polynomials of arbitrary orders and $\bar{\Theta}$ is the affine term. $G_2(q)$'s leading coefficient is zero so that there is at least one sample lag contributing to $y(t)$. The free entries

of $\bar{\Theta}$, F, G_1 and G_2 can be thought of as original model's hyper-parameters. Equation (10) allows a rational dependence of model's parameters on the system's states and inputs. Note that this form of parameter representation makes the model essentially a bilinear structure. Such forms are appealing candidates for modeling many nonlinear processes such as those arising in the areas of fMRI deconvolution and nonlinear tracking. The identification approach that we proposed (see [9] for details) was to first realize the $\Theta(t)$ trajectory of Equation (9) under suitable constraints, followed by the use of the estimated $\Theta(t)$ and the input-output data measurements to estimate the values of $\bar{\Theta}$, $F(q)$, $G_1(q)$ and $G_2(q)$ coefficients. This delivered $\Theta(t)$ expressed as a function of model regressors in a rational form. Under the assumption that F, G_1 and G_2 were *sufficiently sparse*, we treated the contributing regressors as scheduling variables. Briefly, suppose $u(t)$ and $y(t)$ are uniformly sampled and N measurements for $t = 1, 2, \ldots, N$ are available. Consider a state-space realization of the $\Theta(t)$ dynamics in Equation (10):

$$\begin{aligned} X(t+1) &= A_\theta X(t) + B_\theta^1 u(t) + B_\theta^2 y(t) \\ \Theta(t) &= C_\theta X(t) + D_\theta^1 u(t) + \bar{\Theta} \end{aligned} \tag{11}$$

Let $U(t) = [u(t), y(t), 1(t)]^T$ be the augmented input vector of length $p = ny + nu + 1$, where the step input $1(t)$ is added to account for the affine term $\bar{\Theta}$. Then, as described in subspace identification literature, the minimal order of the parameter model (Equation (11)) is equal to the rank of the matrix $H_{n,m,N}(\Theta)H_U^\perp$ where:

$$H_{n,m,N}(\Theta) = \begin{bmatrix} \Theta(1) & \Theta(2) & \ldots & \Theta(N-m+1) \\ \Theta(2) & \Theta(3) & \ldots & \Theta(N-m+2) \\ \vdots & \vdots & \ddots & \vdots \\ \Theta(m) & \Theta(m+1) & \ldots & \Theta(N) \end{bmatrix} \tag{12}$$

$H_U^\perp \in \Re^{(N-m+1)\times q}$ is a matrix whose columns form an orthogonal basis for the null space (nullity q) of the Hankel matrix $H_{p,m,N}(U)$:

$$H_{p,m,N}(U) = \begin{bmatrix} U(1) & U(2) & \ldots & U(N-m+1) \\ U(2) & U(3) & \ldots & U(N-m+2) \\ \vdots & \vdots & \ddots & \vdots \\ U(m) & U(m+1) & \ldots & U(N) \end{bmatrix} \tag{13}$$

Thus, the minimal order Θ dynamics that explains the observed outputs while satisfying some additional sparsity and rate of change constraints. Using nuclear norm as convex relaxation of matrix rank led to the following optimization:

$$\begin{aligned} \underset{\Theta}{\text{minimize}} \quad & \|H_{n,m,N}(\Theta)H_U^\perp\|_* \\ \text{subject to} \quad & \\ & \|y(t) - \Theta(t)^T\Phi(t)\| \leq \delta_1 \\ & plus \quad additional \quad constraints \end{aligned} \tag{14}$$

where the minimization is over the entire $\Theta(t)$ sequence of N samples, δ_1 is a measure of maximum output disturbance. The choice of additional constraints reflects our prior knowledge about the system behavior. For example, we can impose constraints that $\Theta(t)$ changes more slowly than the output of $y(t)$ of the model,

the changes are "smooth", and/or limited to a 1-norm ball with unknown center. These constraints may also be considered as regularizing penalties in the primary objective. Once a trajectory for $\Theta(t)$ is obtained, a linear (affine) model is fit to it using the standard subspace approach.

Bilinear Systems: The proposed representation of $\Theta(t)$ dynamics means that we essentially have a bilinear system with terms composed of lagged input-output variables. As an example, consider the system:

$$A(q)y(t) = B_0 u(t)F(u(t)) + B_1(q)u(t) + e(t) \tag{15}$$

where $A(q)$ and $B_1(q)$ are fixed coefficient polynomials, B_0 is a constant and $F(.)$ is a low-pass filter. This can be expressed in LPV form:

$$A(q)y(t) = B(q,t)u(t) + e(t) \tag{16}$$

where $B(q,t)$ is a time-varying polynomial. For example, if $A(q)$ and $B_1(q)$ are second-order, $B_1(q)$ has no feedthrough term and $F(.)$ is a third-order moving average filter, then $B(q,t) = [b_0(t), b_1 q^{-1}, b_2 q^{-2}]$, $b_0(t) = \sum_{i=0}^{2} \alpha_i u(t-i)$. We then have a second order ARX model with a time-varying input gain $b_0(t)$. The objective function is:

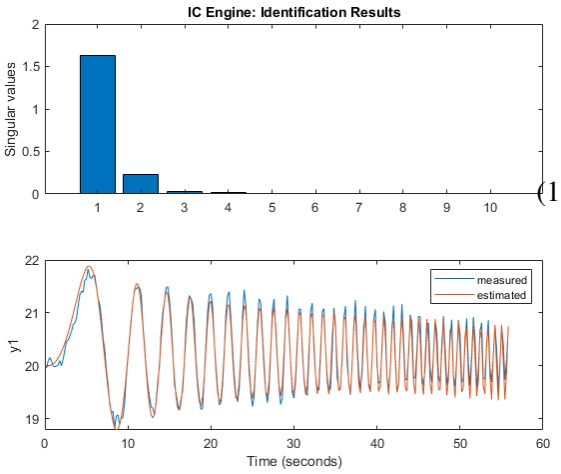


Figure 6: Input-output data. y1 is the output (RPM/100) and u1 is the input (V).

$$\underset{B,\Theta}{\text{minimize}} \|H_{n,m,N}(\Theta)H_U^{\perp}\|_* \text{ subject to } \|y(t) - \Theta^T\Phi(t)\| \leq \delta_1 \text{ and } \|\Delta\Theta_{max}\|_1 \leq \delta_2 \tag{17}$$

The first constraint checks prediction error, the second imposes a limit on the rate of change of parameters $\Theta(t)$. Here $\Delta\Theta_{max}$ is a vector of maximum allowable parameter changes, such that for the i^{th} parameter, the value is $\max_t(|\theta_i(t+1) - \theta_i(t)|)$. The effectiveness of this approach is illustrated in Fig. 6, where it was used to identify the dynamic relationship between the voltage controlling the Bypass Idle Air Valve (BPAV) and the engine speed for an internal combustion engine.

2.1.3 Identifying Dynamical Graphical Models [10]. The research outlined above considered only unstructured models. However, in many practical scenarios unstructured models can lead to models that do not respect the physical constraints of the problem, since they fail to capture the structure of the interactions between physical agents, allowing for non-realistic interactions. Examples of these scenarios range from models of tightly interacting infrastructures (e.g. the power and communication grids) to biological systems and crowd behavior. In order to model these systems, in this portion of the research we developed a novel method for identifying dynamical graphical models, represented by a directed graph structure $G = \{V, E\}$, where each node V corresponds to a given time series (the behavior of a specific agent), and the edges E are operators relating the values of these series at different time instants, accounting for the dynamics arising from agent interactions (see Fig. 7). The corresponding equations are given by

$$\begin{aligned} x_j(t) = \quad & \textstyle\sum_{i=1}^{n} \sum_{k=1}^{r} c_{j,i}(k) x_i(t-k) + \eta_j(t), \\ & t \in [r+1, T],\ j = 1, \ldots, n \end{aligned} \tag{18}$$

where $x_j(.)$ denotes the time series at the j^{th} node, $c_{j,i}(.)$ are the coefficients of an ARX model relating the present value of the time series at node j to the past values measured at node i, and $\eta_j(t)$ represents measurement noise. Our goal was to identify such a structure from experimental data. Note in passing that, unless a regularization criteria is added, the problem is ill posed, since an infinite number of topologies can explain a given set of finite, noisy observations. In this research, we used "sparsity'" to regularize the problem, reflecting the fact that usually the solution with the fewest number of edges is the correct one. Let

$$
\begin{array}{rcl}
\mathbf{x}_j & \doteq & \left[x_j(T), \ldots, x_j(r+1)\right]^T \\
\boldsymbol{\eta}_j & \doteq & \left[\eta_j(T), \ldots, \eta_j(r+1)\right]^T \\
\mathbf{c}_{j,i} & \doteq & \left[c_{j,i}(1), \ldots, c_{j,i}(r)\right]^T \\
\mathbf{c}_j & \doteq & \left[\mathbf{c}_{j,1}^T \ldots, \mathbf{c}_{j,n}^T\right]^T \\
\mathbf{C} & \doteq & \left[\mathbf{c}_1, \ldots, \mathbf{c}_n\right] \\
\mathbf{X} & \doteq & \left[\mathbf{x}_1, \ldots, \mathbf{x}_n\right] \\
\mathbf{H}_i & \doteq & \begin{bmatrix} x_i(T-1) & x_i(T-2) & \ldots & x_i(T-r) \\ x_i(T-2) & x_i(T-3) & \ldots & x_i(T-r-1) \\ \vdots & \ldots & \ldots & \vdots \\ x_i(r) & \ldots & \ldots & x_i(1) \end{bmatrix} \\
\mathbf{H} & \doteq & \left[\mathbf{H}_1 \quad \ldots \quad \mathbf{H}_n\right] \\
\boldsymbol{\Xi} & \doteq & \left[\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n\right]
\end{array}
$$

With this notation, the equations describing the complete model can be written in compact form as:

$$\mathbf{X} = \mathbf{H}\mathbf{C} + \boldsymbol{\Xi} \tag{19}$$

and the problem of interest here can be precisely stated as:

Problem 2. *Given T measurements of n time series $x_i(t),\ i = 1, \ldots, n,\ t \in [1, T]$, and upper bounds ϵ and r on the noise level and edge model order, respectively, solve:*

$$\begin{array}{r} \min \sum_i \|\{\mathbf{c}_i\}\|_0 \text{ s. t. (19) and } \|\boldsymbol{\eta}_i\|_2 \leq \epsilon, \\ \forall i = 1, \ldots, n \end{array} \tag{20}$$

where $\mathbf{c}_i \in \mathbb{R}^r$ and $\|\{\mathbf{c}_i\}\|_0$ denotes the number of non-zero elements of the vector sequence $\mathbf{c}_i$.

Note that the objective function in this problem is precisely $|E|$, the number of edges in the graph, and that, due to its structure, the problem above decouples into n subproblems of the form:

$$\begin{array}{l} \min \|\{\mathbf{c}_i\}\|_0 \text{ s. t. } \|\boldsymbol{\eta}_j\|_2 \leq \epsilon \text{ and} \\ \mathbf{x}_j = \sum_i \mathbf{H}_i \mathbf{c}_i + \boldsymbol{\eta}_j \end{array} \tag{21}$$

This is a (vector) sparsification problem similar to (29) and thus can be solved using a relaxation similar to (30). However, a computationally attractive alternative can be obtained by expanding the concept of atomic norm introduced in §2.1.2 to encompass the case where it is desired to *block*-sparsify a vector sequence. Specifically, given a set of atoms $\mathcal{A} = \{a\}$, assume that it can be partitioned into N centrally symmetric subsets $\mathcal{A}_i$ (the super-atoms), such

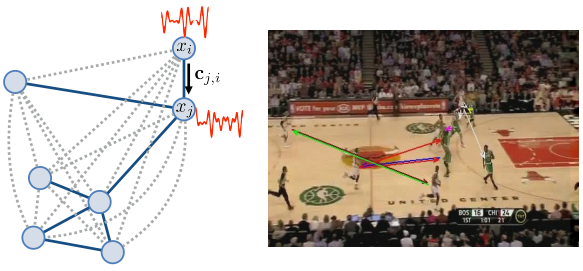


Figure 7: Left: a dynamical graphical model. Right: unveiling interactions between humans by identifying the underlying graphical model.

that $\mathcal{A} = \cup_i \mathcal{A}_i$ and $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset,\ i \neq j$ and associate to each super-atom $\mathcal{A}_i = \{a_{i,1}, ..a_{i,n_i}\}$ the matrix $\mathbf{A}_i$ having as its j$^{\text{th}}$ column $\mathbf{a}_{i,j}$, the coordinates of the atom $a_{i,j}$ in a suitable basis in X. Given a point $\mathbf{x} \in X$, its super-atomic norm is defined as:

$$\|\mathbf{x}\|_{s\mathcal{A}} = \quad \min_{\mathbf{c}} \sum_{i=1}^{N} \|\mathbf{c}_i\|_\infty \text{ s.t } \mathbf{x} = \sum_i \mathbf{A}_i \mathbf{c}_i \tag{22}$$

Since the convex envelope of the cardinality of a vector sequence $\{\mathbf{c}\}$, $\|\mathbf{c}_i\|_\infty \leq 1$ is given by:

$$\|\{\mathbf{c}\}\|_{0,env} = \sum_i \|\mathbf{c}_i\|_\infty$$

it follows that, minimizing the super-atomic norm indeed promotes block-sparsity. Further, problems involving the minimization of a function subject to super-atomic norm constraints can be efficiently solved by using the following modification of Algorithm 1 [10]:

Algorithm 2 Convex minimization subject to super-atomic norm constraints

1: Data: set of super-atoms $\mathcal{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_i, \dots\}$
2: Initialize $\mathbf{x}^{(0)} \leftarrow \tau \mathbf{a}$ for some arbitrary $\mathbf{a} \in \mathcal{A}$
3: **for** k = 0,1,2,3,..., k_{max} **do**
4: $\quad L \leftarrow \arg\min_m \left\{ \min_{\|\mathbf{c}\|_\infty \leq 1} \langle \partial f(\mathbf{x}^{(k)}), \sum \mathbf{a}_{i,m} c_i \rangle \text{ s.t. } \mathbf{a}_{i,m} \in \mathcal{A}_m \right\}$
5: $\quad \mathbf{c} \leftarrow \arg \min_{\|\mathbf{c}\|_\infty \leq 1} \langle \partial f(\mathbf{x}^{(k)}), \sum \mathbf{a}_{i,L} c_i \rangle$ s.t. $\mathbf{a}_{i,L} \in \mathcal{A}_L$.
6: $\quad \mathbf{a} \leftarrow \sum_i \mathbf{a}_{i,L} c_i$
7: $\quad \alpha_k \leftarrow \text{argmin}_{\alpha \in [0,1]} f(\mathbf{x}^{(k)} + \alpha[\tau \mathbf{a} - \mathbf{x}^{(k)}])$
8: $\quad \mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} + \alpha_k[\tau \mathbf{a} - \mathbf{x}^{(k)}]$
9: **end for**

The ideas discussed above can be used to solve Problem 2 by simply defining each super-atom as the collection of columns from the matrices $\mathbf{H}_i$, (e.g a collection of vectors, each containing delayed measurements of the respective time-series):

$$\mathcal{A}_i = \{\mathbf{H}_i(:, t)\},\ t = 1, \dots r$$

leading to a super-atomic norm minimization of the form

$$\min \|\mathbf{z}\|_{s\mathcal{A}} \text{ subject to } \|\mathbf{x}_j - \mathbf{z}\|_2 \leq \epsilon \tag{23}$$

where $\mathbf{z} = \sum_i \mathbf{H}_i \mathbf{c}_i$. Finally, imposing soft, rather than hard constraints on the fitting error leads to:

$$\min \|\mathbf{x}_j - \mathbf{z}\|_2 \text{ subject to } \|\mathbf{z}\|_{s\mathcal{A}} \leq \tau \tag{24}$$

which can be efficiently solved using Algorithm 2. As before, this approach only requires computing inner products and thus can handle large data sets. Further, as shown in [10], it can be easily extended to handle unknown inputs, modeling for instance the interaction of the system with its environment.

2.1.4 Identification of Switched Systems [11–13]. Switched affine systems are important on their own, since they arise in the context of a wide domain of application domains ranging from fault-tolerant control to manufacturing, and as a "poor man's" model of non-linear phenomena. Given their importance, substantial research has been devoted to develop algorithms for stability analysis and controller synthesis for switched systems operating in different scenarios. However, in many practical cases, models of the system under consideration are not available and must be obtained from a combination of experimental data and

a-priori information before these analysis and synthesis tools can be applied. Identification of switched systems has been extensively studied in the past decade, mainly in the context of two different scenarions: (i) error-in-process models and (ii) error-in-variables models. While the first case had been largely solved by the time this research started, the case of error-in-variables models, where input/output measurements are corrupted by noise (and the related output estimation problem where only the outputs are affected by noise) was considerably less developed. Since in this case the problem is known to be NP hard, most existing methods were based upon convex relaxations of the original non-convex problem. However, computational complexity of these relaxations scaled combinatorially both with the number of subsystems and their order, limiting the approach to systems consisting of relatively few low order subsystems.

To address these difficulties, in this portion of the research we developed a method that addressed the computational complexity noted above, while, at the same time providing convergence guarantees in the set theoretic sense. Specifically, we sought to develop a tractable, scalable framework to solve the following Error-in-Variables (EiV) identification problem:

Problem 3. *Given experimental input/output data $\{(u_t, y_t)_{t=t_0}^T\}$ and a bound σ_η on the covariance of the noise, find a set of coefficients $\{a_{k=1}^{n_a}(i), b_{k=1}^{n_b}(i)\}$ so that the EIV-SARX model*

$$\begin{aligned} \hat{y}_t &= \sum_{k=1}^{n_a} a_k(s_t)\hat{y}_{t-k} + \sum_{k=1}^{n_b} b_k(s_t)u_{t-k},\ n_a \geq n_b \\ y_t &= \hat{y}_t + \eta_t \end{aligned} \tag{25}$$

explains the observed data. Here $\hat{y}_t$ represents the actual output, corrupted by additive noise η_t, y_t denotes its measured values, and s_t is the mode variable indicating which subsystem is active at time t.

Our main result was a computationally efficient identification algorithm, based upon the idea of embedding the experimental data in the manifold of positive definite matrices and using a manifold metric to identify time intervals guaranteed to contain no switches. The key idea behind the approach that we developed is to embed data in $\mathcal{S}_{++}^n$, the manifold of positive definite matrices, and use a suitable manifold distance to detect switches and compare systems, exploiting the fact that the manifold distance between data points generated by the same subsystem is substantially smaller than the distance between points corresponding to different subsystems. Thus, switches can be detected by sharp increases in the manifold distance, and segments where the same subsystem is active can be identified by finding clusters where this distance is small, a problem that can be efficiently solved by recasting it into a graph cut form. In particular, in this research we used the Jensen-Bregman Log-Det Divergence (JBLD) given by:

$$J_{ld}(\mathbf{X}, \mathbf{Y}) \doteq \log\left|\frac{\mathbf{X}+\mathbf{Y}}{2}\right| - \frac{1}{2}\log|\mathbf{X}\mathbf{Y}| \tag{26}$$

Attractive properties of the JBLD include the facts that (i) its square root is a geometry aware metric in $\mathcal{S}_{++}^n$, and (ii) its low computational burden, compared against the AIRM. Finally, for a fixed $\mathbf{Y} \in \mathcal{S}_{++}^n$, $J_{ld}(\mathbf{X}, \mathbf{Y})$ is convex in the region $\{\mathbf{X} \in \mathcal{S}_{++}^n : \mathbf{X} \preceq (1+\sqrt{2})\mathbf{Y}\}$. Note that (26) is only well defined for matrices in $\mathcal{S}_{++}^n$ while our research required comparing positive semi-definite matrices. The following result extending the JBLD to $\mathcal{S}_{+}^n$, provides the theoretical justification for the proposed identification method (see [13] for a proof).

Theorem 2 ([11])**.** *Given $\mathbf{X}, \mathbf{Y} \in \mathcal{S}_{+}^n$, define the regularized matrices $\mathbf{X}(\sigma) = \mathbf{X} + \sigma\mathbf{I},\ \mathbf{Y}(\sigma) = \mathbf{Y} + \sigma\mathbf{I}$, where $\sigma > 0$. Then*

$$\lim_{\sigma \to 0} J_{ld}(\mathbf{X}(\sigma), \mathbf{Y}(\sigma)) \neq \infty \iff \mathcal{N}(\mathbf{X}) = \mathcal{N}(\mathbf{Y}) \tag{27}$$

Corollary 1. *Consider two time series* $\mathbf{x_k}$ *and* $\mathbf{y}_k$*. Let* $\mathbf{H_x}, \mathbf{H_y}$ *denote the corresponding Hankel matrices and define* $\mathbf{G_x}(\sigma) \doteq \mathbf{H_x}\mathbf{H_x}^T + \sigma\mathbf{I}$ *and* $\mathbf{G_y}(\sigma) \doteq \mathbf{H_y}\mathbf{H_y}^T + \sigma\mathbf{I}$*. Then the time series* $\mathbf{x}_k$ *and* $\mathbf{y}_k$ *are behaviors of the same underlying dynamics if and only if* $\lim_{\sigma\to 0} J_{ld}(\mathbf{G_x}(\sigma), \mathbf{G_y}(\sigma)) \neq \infty$

Direct application of this corollary led to the Algorithm outlined below:

Algorithm 3 Riemannian distance-like function based switched system identification

Inputs: input sequence $\mathbf{u}_{1:n}$, output sequence $\mathbf{y}_{1:n}$, sliding window size h, $r = n_a + n_b + 1$.
Step 1: Data Segmentation. Use Corollary 1 to partition the input and output sequences into segments of length h, each generated by a single LTI system. The ith segment of input and output are denoted $\mathbf{u}_{i:i+h-1}$ and $\mathbf{y}_{i:i+h-1}$, respectively.
Step 2: Spectral Clustering.
for i = 1 to # of segments **do**
$\mathbf{H}_{\mathbf{u},i} \leftarrow$ Hankelize $\mathbf{u}_{i:i+h-2}$
$\mathbf{H}_{\mathbf{y},i} \leftarrow$ Hankelize $\mathbf{y}_{i:i+h-1}$
$\mathbf{G}_i \leftarrow \begin{bmatrix}\mathbf{H}_{\mathbf{y},i}\\ \mathbf{H}_{\mathbf{u},i}\end{bmatrix}\begin{bmatrix}\mathbf{H}_{\mathbf{y},i}^T & \mathbf{H}_{\mathbf{u},i}^T\end{bmatrix}$
$\hat{\mathbf{G}}_i(\sigma) \leftarrow \frac{\mathbf{G}_i}{\|\mathbf{G}_i\|_*} + \sigma\mathbf{I}_r$
end for
Compute $\mathbf{D}$, where $\mathbf{D}_{ij} = J_{ld}(\hat{\mathbf{G}}_i(\sigma), \hat{\mathbf{G}}_j(\sigma))$
Compute the similarity matrix $\mathbf{W}$, where $\mathbf{W}_{ij} = e^{-\frac{D_{ij}^2}{2}}$
Cluster labels $\mathbf{z} \leftarrow$ normalized cuts on $\mathbf{W}$
Step 3: Subsystem Identification.
for i = 1 to # of clusters **do**
Perform a LTI SysId step on cluster z_i
end for
Outputs: $\mathbf{a}_i, \mathbf{b}_i$

The main result of this portion of the research showed that the algorithm is guaranteed to recover the correct data segmentation and model parameters if the noise is below a certain threshold that depends only on the problem data. Further, consistent numerical experience has shown that the algorithm performs well even when for noise levels above this threshold [13]. Finally, note that the segmentation step scales *linearly* with the number of data points, while the spectral clustering step scales as (number of clusters)^3. Thus, the algorithm has moderate computational complexity, even for large data sets, provided that the number of switches is not too large.

2.1.5 Identification and decision making in the presence of gross outliers [14–17]. So far, we have considered scenarios with noisy data, but where all this data has been generated by the system to be identified. However, many practical scenarios involve situations where the data is corrupted by outliers that, if not corrected, can severely affect the quality of decisions taken based on the data. Examples of these situations include control decisions in the face of sensor outages or tracking using data corrupted while transmitted over a wireless link. Thus, a prerequisite to provably robust decision making in realistic scenarios is the ability to find a model that interpolates the largest number of data points (the "inliers"), while, at the same time, identifying the outliers. The simplest version of this problem is given by:

Problem 4. *Given noisy input/output data* $\{(u_t, y_t)_{t=t_0}^T\}$*, find an ARX model of the form*

$$y_t = \sum_{k=1}^{n_a} a_k y_{t-k} + \sum_{k=1}^{n_b} b_k u_{t-k} + \eta_t \tag{28}$$

that maximizes the number of points interpolated.

It is well known that even this simplified problem is generically NP hard and scales combinatorially with the number of data points. However, as we have shown, efficient convex relaxations, in many cases with optimality certificates, can be obtained by recasting the problem into either (i) a regularized robust regression form or (ii) a polynomial optimization form.

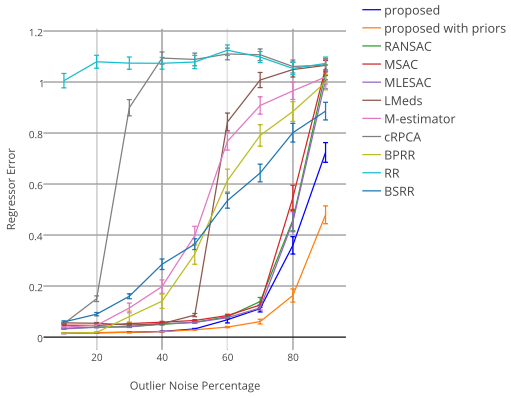


Figure 8: Identification error as a function of the number of outliers for a synthetic 5^{th} order system. Note that by exploiting priors, our method can handle up to 80% outliers.

(i) Identification with outliers as a robust regression problem [14]:
Define $\mathbf{r} = [a_1..a_{n_a}\ b_1..b_{n_b}]^T$ and $\mathbf{x}_t = \begin{bmatrix} y_{t-1} & \cdots & y_{t-n_a} & u_{t-1} & \cdots & u_{t-n_b} \end{bmatrix}^T$. In this context, Problem 4 can be compactly stated as finding a parameter vector $\mathbf{r}$ that maximizes the cardinality of the set $\mathcal{T} \doteq \{t : \left|y_t - \mathbf{r}^T\mathbf{x}_t\right| \leq \epsilon\}$. As shown in [14], by introducing additional variables $\mathbf{r}_i \in \mathbb{R}^d$ the problem can be reformulated as:

$$\begin{aligned} \mathbf{r}^* \quad = \quad & argmin_{\mathbf{r},\mathbf{r}_i} \|\{\mathbf{r} - \mathbf{r}_i\}\|_0 \text{ subject to:} \\ & |y_i - \mathbf{x}_i^T\mathbf{r}_i| \leq \epsilon,\ i = 1, \ldots, N \end{aligned} \tag{29}$$

where $\|\{\mathbf{r} - \mathbf{r}_i\}\|_0$ denotes the cardinality (e.g. number of non-zero vectors) of the sequence $\{\mathbf{r} - \mathbf{r}_i\}_{i=1}^N$. While this problem is still generically NP hard, we have proved [14] that its tightest convex relaxation is given by:

$$\begin{aligned} & \mathbf{r}_{env} = argmin_{\mathbf{r},\mathbf{r}_i} \textstyle\sum_{i=1}^N \|\mathbf{r} - \mathbf{r}_i\|_\infty \text{ subject to:} \\ & |y_i - \mathbf{x}_i^T\mathbf{r}_i| \leq \epsilon,\ i = 1, \ldots, N \end{aligned} \tag{30}$$

leading to convex problems that can be efficiently solved. As we proved in [14], under suitable conditions reminiscent of the restricted isometry property, this approach is guaranteed to recover the sparsest solution to (29). Further, it has substantially better robustness properties than existing approaches, due to a "built-in" self-scaling property, where the data is automatically scaled to mitigate the effects of gross outliers. This is illustrated in Fig. 8 showing that, in cases where the outlier spaces are well separated from the data, the proposed method can handle up to 80% of outliers, substantially outperforming existing methods.

(ii) A polynomial optimization based approach [15]:
As noted above, under certain conditions, Problem 4 can be exactly solved by the convex relaxation (30). However, when these conditions fail, $\mathbf{r}_{env}$ may provide a poor approximation to $\mathbf{r}^*$. To avoid this difficulty, we developed an alternative approach based on polynomial optimization. This approach is guaranteed to recover the optimal solution $\mathbf{r}^*$, at the price of increased computational complexity. The idea is to associate

to each pair $(y_i, \mathbf{x}_i)$ a binary variable $s_i \in \{0, 1\}$ that indicates whether the point is an inlier ($s_i = 1$) or outlier ($s_i = 0$), allowing for recasting Problem 4 into the following polynomial optimization form:

$$\begin{aligned} p^* = \max_{s_j, \mathbf{r}} \quad & \textstyle\sum_{j=1}^{N_p} s_j \\ \text{s.t.} \quad & |s_j(y_j - \mathbf{r}^T \mathbf{x}_j)| \leq \epsilon s_j,\ s_j^2 = s_j,\ \forall_{j=1}^{N_p} \\ & \mathbf{r}^T \mathbf{r} = 1,\ \mathbf{r}(1) \geq 0 \end{aligned} \tag{31}$$

Here, the first constraint enforces that $(y_j, \mathbf{x}_j)$ is an inlier when $s_j \neq 0$ and is trivially satisfied otherwise; the second is simply a restatement of the fact that $s_j \in \{0, 1\}$, while the third and fourth constraints normalize the vector $\mathbf{r}$ and remove ambiguities. Clearly, since $s_i = 1 \iff (y_i, \mathbf{x}_i)$ is an inlier, the objective maximizes the number of inliers.

Problem (31) is a polynomial optimization problem and thus can be solved using sum-of-squares (SoS) and moments techniques to reduce it to a convex semi-definite program [15, 18]. Further as shown in [15], this problem exhibits the *running intersection property* (e.g. the underlying correlative sparsity graph is sparse), allowing for developing an efficient algorithm whose complexity scales linearly with the number of data points and can handle co-occurrence priors. An example of this situation is enforcing that a given set of points are either all inliers or outliners, a situation arising for instance when the minimum length of sensor outages is known, or when separating background from foreground.

2.1.6 Semi-supervised identification of switched systems in the presence of outliers [15–17]. In the initial phase of this research, we have addressed identification of time-invariant systems. Subsequently, we sought to extended the framework to switching systems. These systems are interesting in their own, since they appear in many scenarios (biological systems transitioning amongst different metabolic stages, human activity, physical systems with different operation modes, etc) and as tractable approximations to more complex non-linear dynamics. For simplicity, we consider only single-input single-output systems, but extension to the MIMO case is straightforward. Specifically, we considered the following extension of Problem 4:

Problem 5. *Given:*

- *A set of input/output data $\{(u_t, y_t)_{t=t_0}^T\}$ generated by an SARX model of the form*

$$y_t = \sum_{k=1}^{n_a} a_k(\sigma_t) y_{t-k} + \sum_{k=1}^{n_b} b_k(\sigma_t) u_{t-k} + \eta_t \tag{32}$$

- *A-priori information consisting of (i) a bound N_s on the number of subsystems (e.g. $\sigma_t \in \mathbb{N}_{N_s}$), (ii) a bound ϵ on the process noise η_t, (iii) additional information, such as N_{f_i}, the relative frequency of each submodel, point wise co-occurrences, constraints on the switching sequence, etc.*

Find a set of coefficients $\{a_{k=1}^{n_a}(i), b_{k=1}^{n_b}(i)\}$, each associated with the submodel $G_i, \forall_{i=1}^{N_s}$, that maximizes the number of inliers.

As in section 2.1.5, we pursued two alternative approaches to solving the problem above, one based on robust-regression and the second based on polynomial optimization.

(i) A sparsification based approach [15]: The idea underlying this approach is to find one submodel at a time, by successively finding a parameter vector $\mathbf{r}$ that makes $\left|y_t - \mathbf{x}_t^T \mathbf{r}\right| \leq \epsilon$ feasible for as many time

instants t as possible. Once this model is found, the points explained by it are removed from the data set and the procedure is repeated until all data points are clustered. By considering at each stage, points not explained by the model are treated as outliers, each parameter vector $\mathbf{r}_i$ can be found using the algorithm developed in §2.1.3. When the subspaces spanned by each subsystem are well separated, the recovery results in [14] guarantee that this approach will indeed find the correct set of models. On the other hand, if these conditions do not hold, due to its greedy nature, the algorithm can overestimate the number of subsystems required to explained the observed data. Nevertheless, consistent numerical experience shows that these instances are very rare, specially when the algorithm is combined with a re-weighted heuristics to enhance sparsity.

(ii) A polynomial optimization based approach [15]: As in §2.1.3, an alternative approach, with optimality certificates can be obtained by recasting the problem into a polynomial optimization form, by introducing a set of binary variables $s_{i,t}$ that indicate whether the submodel G_i is active at time instant t ($\sigma_t = i \Leftrightarrow s_{i,t} = 1$) and auxiliary variables $e_{i,t}$ (the fitting error of point t to model i), leading to:

$$\begin{array}{l} \max\limits_{\mathbf{r}_i, e_{i,t}, s_{i,t}} \quad \sum_{t=t_o}^{T} \sum_{i=1}^{N_s} s_{i,t} \\ \text{subject to} \\ \mathbf{r}_i^T \mathbf{x}_t - e_{i,t} = 0 \\ ||\mathbf{r}_i||_2 = 1 \\ |s_{i,t} e_{i,t}| \leq \epsilon s_{i,t} \\ s_{i,t}^2 = s_{i,t}, \ \sum_{i=1}^{N_s} s_{i,t} \leq 1 \\ \forall i = 1, \cdots, N_s, \forall t = t_o + n_a, \cdots, T \end{array} \tag{33}$$

Here, $\mathbf{r}_i(j)$ is the j^{th} entry of $\mathbf{r}_i$, and the last two constraints on $s_{i,t}$ guarantee that no more than one submodel is active at time instant t. As before, it can be easily shown that this problem satisfies the running intersection property, leading to algorithms that scale linearly with the number of data points.

In many practical scenarios, additional information is available about the system to be identified. Examples of these situations include knowledge that certain transitions are inhibited (common in biological applications), or co-occurrences (common in image processing and computer vision) where some of the data may be manually annotated, so that it is known that two given data points belong (or do not belong) to the same system. A salient feature of the approach above is its ability to incorporate these priors by imply imposing additional constraints on the variables $s_{i,t}$. For instance:

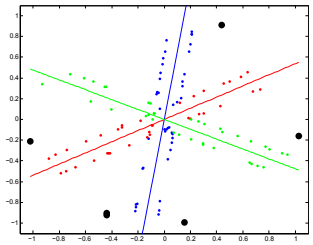


Figure 9: A switching system with 3 subsystems. The polynomial based approach correctly identified both the outliers (black dots) and the parameters of each subsystem.

(i) submodel G_i is active for $f\%$ of the time $\Longleftrightarrow \sum_{t=t_0+n_a}^{T} s_{i,t} = 0.01 f (T + 1 - t_0 - n_a)$;

(ii) the same submodel is active at time instants m and $n \Longleftrightarrow s_{i,m} = s_{i,n}$, $\forall i = 1, \cdots, N_s$;

(iii) different submodels are active at time instants m and $n \Longleftrightarrow s_{i,m} s_{i,n} = 0$, $\forall i = 1, \cdots, N_s$;

(iv) submodel i cannot be followed by submodel $j \Longleftrightarrow s_{i,t} s_{j,t+1} = 0, \forall t$.

(iii) A solution based on Christoffel Polynomials [16, 17]: While the approach outlined above is guaranteed to find the correct system, its computational complexity is very high, even when exploiting the underlying sparse structure of the problem. As an alternative, we developed a computationally cheaper method, with stochastic certificates, by exploiting the connections between positive Borel measures, positive polynomials and Christoffel functions. Specifically, we sought to solve the following problem:

Problem 6. *[Robust Subspace Clustering] Given a (sufficiently dense) set of noisy i.i.d samples (with arbitrary, unknown distribution μ) $\mathbf{x}_i$ of points $\hat{\mathbf{x}}_i \in \mathbb{R}^d$ drawn from an arrangement of (linear) subspaces $\mathcal{A} \doteq S_1 \cup S_2 \cup \ldots S_n$, corrupted with gross outliers, and a bound n on the number of subspaces: 1) Identify gross outliers in the data; and 2) Assign inliers to subspaces.*

The main idea of the algorithm that we proposed, illustrated in Fig. 10, is to find a polynomial $P_s(\mathbf{x})$ whose level sets approximate the support of the inliers (that is, the polynomial is small in the inlier subspaces and has large values elsewhere). This polynomial is obtained by (i) first identifying "reliable" inliers, (ii) estimating the probability density function of these inliers, and (iii) finding a polynomial that minimizes the expected value of the quadratic fitting error to the (estimated) inlier subspaces. Once this polynomial approximation to the support set of the inliers is available, true inliers can be found by simply seeking points where $P_s(\mathbf{x})$ is small. Notably, as discussed in the sequel, steps (i)-(iii) only involve performing two singular value decompositions on a matrix whose size is, in typical applications, much smaller than the data matrix, and thus are computationally far less costly than solving a regularized optimization problem.

In order to explain our results, we need to introduce first the following background results on positive polynomials and Borel measures:

Moment matrices and orthogonal polynomials Given a probability measure μ supported on $\mathbb{R}^d$, its corresponding moments sequence is given by

$$m_{\boldsymbol{\alpha}} = \mathcal{E}_\mu(\mathbf{x}^{\boldsymbol{\alpha}}) = \int_{\mathbb{R}^d} \mathbf{x}^{\boldsymbol{\alpha}} d\mu \tag{34}$$

where $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_d \end{bmatrix}$ is a multi-index , $\mathbf{x} \doteq \begin{bmatrix} x_1 & x_2 & \ldots & x_d \end{bmatrix}^T$ and $\mathbf{x}^{\boldsymbol{\alpha}} \doteq x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$.

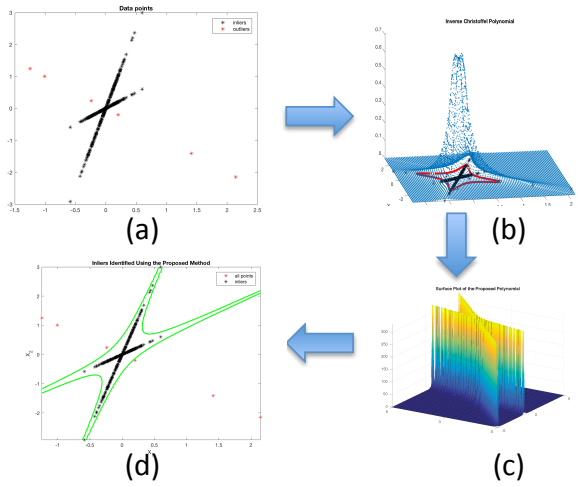


Figure 10: Proposed pipeline: (a) inliers (black) corrupted with outliers (red); (b) SoS approximation of the support of the inliers (blue) and its level set (red), (c) approximating the support of the inlier subspace using the inverse of the proposed polynomial; (d) inliers are the points inside a suitable level set of the polynomial.

In the sequel, we are interested in using a matrix representation of a given sequence $\mathbf{m}$ that contains all the moments up to order $2n$. To this effect, we will arrange the moments according to a graded reverse lexicographic order (grevlex) of the corresponding monomials and form the matrix $\mathbf{M}_n \in \mathbb{R}^{s_{n,d} \times s_{n,d}}$ with entries given by $\mathbf{M}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = m_{\boldsymbol{\alpha}+\boldsymbol{\beta}}$. For example, for $d = 2$ variables, x_1 and x_2, there are $s_{n,d} = 6$ monomials of degree up to $n = 2$: $1, x_1, x_2, x_1^2, x_1 x_2, x_2^2$. The elements of the corresponding moment matrix $\mathbf{M}_2$ with all moments of order up to $2n = 4$, are given by: $m_{(i,j)} = \mathcal{E}_\mu(x_1^i . x_2^j)$, with $0 \leq i, j \leq 3$.

Given a point $\mathbf{x} \in \mathbb{R}^d$, let

$$\boldsymbol{\phi}_n(\mathbf{x}) \doteq \begin{bmatrix} 1 & x_1 & x_2 & \dots (x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}) \dots x_d^n \end{bmatrix}^T \tag{35}$$

where $\sum_{i=1}^d \alpha_i \leq n$. It can be easily seen that

$$\mathbf{M}_n = \int_{\mathbb{R}^d} \boldsymbol{\phi}_n(\mathbf{x}) \boldsymbol{\phi}_n^T(\mathbf{x}) d\mu \tag{36}$$

Thus, $\mathbf{M}_n \succeq 0$ (in fact $\mathbf{M}_n \succ 0$, except when $\mathbf{x}$ belongs to the zero set of a polynomial of degree at most n). The measure μ induces an inner product in $\mathcal{P}_d^n$, the space of polynomials in d variables of degree at most n, given by:

$$\langle P_1(.), P_2(.) \rangle_\mu \doteq \int_{\mathbb{R}^d} P_1(\mathbf{x}) P_2(\mathbf{x}) d\mu \tag{37}$$

As shown next, this inner product can be computed directly from $\mathbf{M}_n$. Consider a generic polynomial $P(\mathbf{x}) \in \mathcal{P}_d^n$, $P(\mathbf{x}) = \sum_\alpha p_\alpha \mathbf{x}^\alpha$, where p_α denotes the coefficients of $P(.)$ in the canonical monomial basis. Collecting all the coefficients p_α in a vector $\mathbf{p}$ allows for compactly representing $P(.)$ as $P(\mathbf{x}) = \mathbf{p}^T \boldsymbol{\phi}_n(\mathbf{x})$. Then,

$$\langle P_1(.), P_2(.) \rangle_\mu = \int_{\mathbb{R}^d} \mathbf{p}_1^T \boldsymbol{\phi}_n(\mathbf{x}) \boldsymbol{\phi}_n^T(\mathbf{x}) \mathbf{p}_2 d\mu = \mathbf{p}_1^T \mathbf{M}_n \mathbf{p}_2 \tag{38}$$

Assume that $\mathbf{M}_n \succ 0$ and let $\mathbf{u}_i, \sigma_i$ denote its singular vectors and the corresponding singular values. From the derivation above, it follows that the polynomials associated with the coefficient vectors $\mathbf{c}_i \doteq \frac{1}{\sqrt{\sigma_i}} \mathbf{u}_i$ are an orthonormal basis, with respect to μ, of $\mathcal{P}_d^n$. These orthonormal polynomials define a reproducing Kernel:

$$K_n(\mathbf{x}, \mathbf{y}) \doteq \sum_{i=1}^{s_{n,d}} (\mathbf{c}_i^T \boldsymbol{\phi}_n(\mathbf{x}))(\mathbf{c}_i^T \boldsymbol{\phi}_n(\mathbf{y})) \tag{39}$$

Next, define the polynomial

$$Q_n(\mathbf{x}) \doteq K_n(\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{s_{n,d}} (\mathbf{c}_i^T \boldsymbol{\phi}_n(\mathbf{x}))^2 \tag{40}$$

Note that $Q_n(\mathbf{x})$ is a sum-of-squares polynomial, and hence non-negative (in fact, it can be shown that $Q_n(\mathbf{x}) \geq 1$). The function $Q_n(\mathbf{x})^{-1}$ is known as the Christoffel function and it is related to the probability measure μ that induces orthorgonality of the set $\{\mathbf{c}_i\}$ through the following result:

$$Q_n(\mathbf{x})^{-1} = \min_{P \in \mathcal{P}_d^n} \int_{\mathbb{R}^p} P^2(\boldsymbol{\xi}) d\mu \text{ s. t. } P(\mathbf{x}) = 1 \tag{41}$$

a result that will play a key role in identifying outliers.

Finding reliable inliers: To find reliable inliers, we started by considering the matrix $\mathbf{M}_{n,h} \doteq \frac{1}{N_p} \mathbf{V}_n^T \mathbf{V}_n \in \mathbb{R}^{s_{n,d-1} \times s_{n,d-1}}$. As in (37), $\mathbf{M}_{n,h}$ induces an inner product in $\mathcal{P}_{d,h}^n$, the space of homogeneous polynomials, given by:

$$\langle P_1(.), P_2(.) \rangle_\mu \doteq \int_{\mathbb{R}^d} P_1(\mathbf{x}) P_2(\mathbf{x}) d\mu = \mathbf{p}_1^T \mathbf{M}_{n,h} \mathbf{p}_2 \tag{42}$$

where $P_i(\mathbf{x}) = \mathbf{p}_i^T \boldsymbol{\nu}_n(\mathbf{x})$. Next, let $\mathbf{U}\mathbf{diag}(\sigma_i)\mathbf{U}^T = \text{svd}(\mathbf{M}_{n,h})$ and consider the homogeneous polynomials $C_i(\mathbf{x}) \in \mathcal{P}_{n,h}$, $C_i(\mathbf{x}) \doteq \mathbf{c}_i^T \boldsymbol{\nu}_n(\mathbf{x})$ with coefficient vectors $\mathbf{c}_i \doteq \frac{1}{\sqrt{\sigma_i}}\mathbf{u}_i$, where $\mathbf{u}_i$ denotes the i^{th} column of $\mathbf{U}$. Note that these polynomials form an orthonormal basis of $\mathcal{P}_{n,h}$ with respect to the inner product defined by (42) since

$$\langle C_i(.), C_j(.)\rangle_\mu = \frac{1}{\sqrt{\sigma_i \sigma_j}} \mathbf{u}_i^T \mathbf{M}_{n,h} \mathbf{u}_j = \begin{cases} 1 & \text{if i=j} \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$K_{n,h}(\mathbf{x}, \mathbf{y}) \doteq \sum_{i=1}^{s_{n,d-1}} (\mathbf{c}_i^T \boldsymbol{\nu}_n(\mathbf{x}))(\mathbf{c}_i^T \boldsymbol{\nu}_n(\mathbf{y})) \tag{43}$$

is a reproducing kernel in this space, with associated Christoffel polynomial given by

$$Q_{n,h}(\mathbf{x}) \doteq K_{n,h}(\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{s_{n,d-1}} (\mathbf{c}_i^T \boldsymbol{\nu}_n(\mathbf{x}))^2 \tag{44}$$

Due to orthonormality, $\mathcal{E}_\mu[C_i(\mathbf{x})] = \mathbf{c}_i^T \mathbf{M}_{n,h} \mathbf{c}_i = 1$. Thus, the expected value of $Q_{n,h}$ is given by

$$\mathcal{E}_\mu(Q_{n,h}) = \sum_{i=1}^{s_{n,d-1}} \mathcal{E}_\mu[c_i(\mathbf{x})] = s_{n,d-1} \tag{45}$$

Finally, given a threshold t, from Markov's inequality it follows that

$$\text{prob}\{Q_{n,h}(\mathbf{x}) \geq t \cdot s_{n,d}\} \leq \frac{1}{t} \tag{46}$$

Hence, a set of "reliable" inliers can be found by simply selecting those points where $Q_{n,h}(\mathbf{x}) < t \cdot s_{n,d}$. Since accurate estimation of a polynomial in the vanishing ideal $I(\mathcal{A})$ hinges on using a set without outliers, t in this step should be taken reasonably low (e.g. $t \sim 1 - 3$).

Estimating a polynomial that vanishes on all inlier subspaces: Since $K_{n,h}$ defined in (43) is a reproducing kernel in $\mathcal{P}_{n,h}$, it follows that, for each point $\mathbf{x} \in \mathbb{R}^d$

$$Q_{n,h}(\mathbf{x})^{-1} = \min_{P \in \mathcal{P}_{d,h}^n} \int_{\mathbb{R}^p} P^2(\boldsymbol{\xi}) d\mu \text{ s. t. } P(\mathbf{x}) = 1 \tag{47}$$

For a given $\mathbf{x}^*$, denote by P^* the minimizer above. Markov's inequality implies that, for any given threshold t, the mass of the set $\mathcal{S} = \{\mathbf{x} \colon (P^*(\mathbf{x}))^2 \leq t\}$ satisfies

$$\begin{aligned} \mu(\mathcal{S}) &= \int_{P^*(\mathbf{x})^2 \leq t} d\mu \geq 1 - \frac{\mathcal{E}_\mu[(P^*)^2)]}{t} \\ &= 1 - \frac{1}{t Q_{n,h}(\mathbf{x}^*)} \geq 1 - \frac{\sigma_{min}}{t[\mathbf{u}_{min}^T \boldsymbol{\nu}(\mathbf{x}_o)]^2} \end{aligned} \tag{48}$$

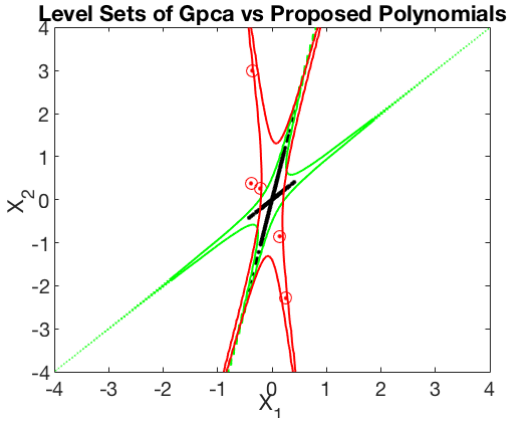


Figure 11: A few outliers (red points) can lead to a GPCA polynomial whose level sets (red) are a poor approximation of the inlier subspaces while the level sets of the proposed polynomial (green) successfully approximate the support of these subspaces.

where σ_{min} denotes the minimum singular value of the moments matrix. It follows that, when $[\mathbf{u}_{min}^T \boldsymbol{\nu}(\mathbf{x}_o)]^2 \gg \sigma_{min}$ (roughly speaking the outlier is well outside the noise level) then the set $\mathcal{S}$ contains most of the inliers (see [17] for details). The effectiveness of this approach in capturing the distribution of the true inliers and in detecting outliers is illustrated in Figure 11.

2.1.7 Optimal Identification and Filtering on the Positive Definite Manifold [19, 20]. In principle, computationally efficient solutions to a large class of problems related to robust inferencing can be obtained by encapsulating all the relevant information in a positive definite matrix (for instance containing the moments of the probability distribution characterizing the process). Examples range from target tracking and re-identification to activity recognition and fault detection. However, successful implementation of these ideas requires the ability to estimate a time-varying positive definite matrix from a collection of noisy measurements, with the main difficulty stemming from the fact that, compared against traditional parameter estimation methods, here the estimation algorithm should respect the fact that the matrix evolves on the PD manifold. Specifically, the problem of interest can be formally stated as:

Problem 7. *Given a noisy observation $\mathbf{Q}_t$ of a positive semi-definite matrix $\mathbf{P}_t$ and past historical data $\mathbf{P}_{t-r}, \mathbf{P}_{t-r+1}, \cdots, \mathbf{P}_{t-1}$, find the Jensen Bregman LogDet (JBLD) divergence-based maximum likelihood estimate (MLE) of $\mathbf{P}_t$. where, for two positive definite matrices $\mathbf{X}, \mathbf{Y}$, the JBLD[1] was defined in* (26) *(recall that the JBLD is a metric that respects the Riemannian topology of the positive definite manifold).*

To solve the problem above, we developed a new framework for recursive filtering on the PD manifold using Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models for propagating past measurements, combined with a maximum likelihood estimator based on minimizing the Jensen Bregman LogDet (JBLD) divergence. Specifically, we introduced a new probabilistic dynamic model for recursive filtering on the PD manifold based on a generalized Gaussian distribution. As shown in [19, 20], under suitable conditions, the generalized Gaussian conjugate prior can indeed be expressed in terms of the JBLD distance between the observed and predicted data, leading to the following model for estimating $\mathbf{P}_t$ based on past observations:.

$$p(\mathbf{P}_t|\mathbf{P}_{t-1}, \ldots, \mathbf{P}_{t-r}) \propto e^{-\frac{J_{ld}(\mathbf{P}_t, \sum_{i=1}^r \mathbf{S}_{t-i}\mathbf{A}_i\mathbf{S}_{t-i})}{2\omega^2}} \tag{49}$$

where $\mathbf{S}_{t-i} \doteq \mathbf{P}_{t-i}^{\frac{1}{2}}$, r denotes the system order and where $\mathbf{A}_i \succ 0$ are the parameters that define the autoregressive model. Intuitively, the probability of obtaining a given covariance at time t decays exponentially with its distance, measured in the J_{ld} sense, from the predictions of the model $\sum_{i=1}^r \mathbf{S}_{t-i}\mathbf{A}_i\mathbf{S}_{t-i}$. In this context, Problem 7 above can be split into the following two subproblems:

Problem 8. *Given a sequence of training data $\{\mathbf{P}_t\}_{t=1}^T \in \mathbb{S}_{++}^n$, find the JBLD-based maximum a posteriori estimate (MAP) of the parameters $\mathbf{A}_i$, such that the dynamic model is stable.*

Problem 9. *Given a noisy observation $\mathbf{Q}_t$, find the JBLD-based maximum likelihood estimate of $\mathbf{P}_t$, assuming a known propagation model of the form* (49).

As shown in [19, 20], the first problem is convex and can be efficiently solved using an ADMM type algorithm while the second admits a closed form solution, leading to a framework for data driven prediction in the positive definite manifold that outperforms existing approaches both in quality of the predictions and computational complexity (Fig. 12).

[1] Recall that the JBLD is a metric that respects the Riemannian topology of the positive definite manifold.

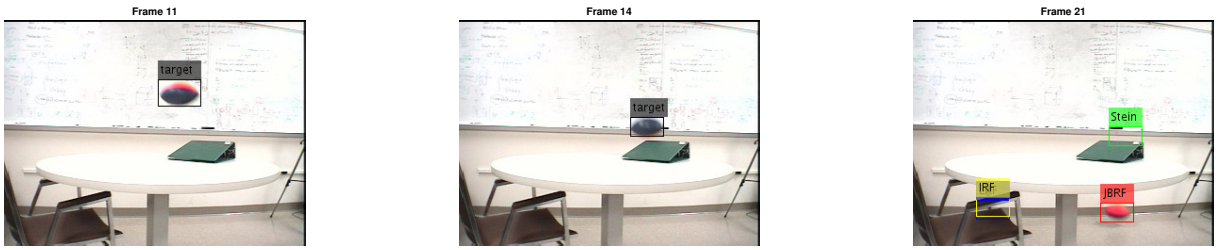


Figure 12: Tracking a target with changing appearance via positive definite embeddings. A GARCH model was used to predict the appearance of the target and locate it. Note that the JBLD based recursive filter (JBRF) is the only method capable of tracking the target through the occlusion [20].

2.2 Model (In)Validation [5, 7].

Model (in)validation is the dual problem of identification: given a model and experimental data, the goal here is to determine whether these are consistent. That is, whether or not the observed data (corrupted by noise) could have been generated by the model. Validating identified models against additional data is a key step before using these models for control synthesis. Additional applications of model (in)validation include fault detection and isolation, and, interestingly, to detect anomalies in time series, including abnormal human activity. The original (in)validation framework developed in the late 1990's and early 2000's considered LTI systems described by a single model. While this was a good start, realistic scenarios typically involve systems that switch amongst different models (e.g. an aircraft switching amongst different regimes, a biological system transitioning between metabolic stages, or a human performing an activity composed of different sub-activities). Thus, as part of this research we developed a framework for (in)validating switched models for cases where the mode variable is not directly accessible (hence, it is not known which submodel generated a specific observation). Formally, the problem of interest here can be posed as determining whether a noisy input/output sequence could have been generated by a given model of the form:

$$\begin{array}{rcl} \xi_t & = & \sum_{i=1}^{n_a} a_i(\sigma_t)\xi_{t-i} + \sum_{i=1}^{n_b} b_i(\sigma_t)u_{t-i} \\ y_t & = & \xi_t + \eta_t,\ \sigma_t \in \{1,\ldots,s\},\ \|\eta_t\|_\infty \leq \epsilon \end{array} \tag{50}$$

where y_t denotes the measured output corrupted by the noise η_t. As in the identification case, this problem is known to be generically NP-hard, due to the presence of noise and because the mode variable σ_t is not directly measurable. However, it is possible to obtain tractable relaxations by using sparsification and polynomial optimization tools. This can be done by noting that (50) holds if and only if there exist a set of "indicator" variables $s_{i,t}$ and admissible noise sequence η_t such that

$$\begin{array}{l} s_{i,t}\left(\mathbf{g_{i,t}} - \mathbf{h_i}\boldsymbol{\eta}_{t:t-n_a}\right) = \mathbf{0}\ \forall\ t \in [t_o + n_a, T] \\ \text{subject to} \\ \sum_{i=1}^{n_s} s_{i,t} = 1 \\ s_{i,t} \in \{0,1\} \text{ and } \|\boldsymbol{\eta}_t\|_\infty \leq \epsilon \end{array} \tag{51}$$

where the notation was simplified by defining:

$$\begin{array}{l} \mathbf{g_{i,t}} \doteq a_1(i)y_{t-1} + .. + a_{n_a}(i)y_{t+n_a} \\ -y_t + b_1(i)u_{t-1} + \ldots + b_{n_b}(i)u_{t-n_c} \\ \mathbf{h_i} \doteq \begin{bmatrix} -1 & a_1(i) & \ldots & a_{n_a}(i) \end{bmatrix} \\ \boldsymbol{\eta}^T_{t:t-n_a} \doteq \begin{bmatrix} \eta_t^T & \ldots & \eta_{t-n_a}^T \end{bmatrix}^T \end{array}$$

As shown in [5,7], the (generically non-convex) problem above can be solved using either (i) a sparsification based approach where the variables $s_{i,t}$ are relaxed to be real, rather than binary, subject to a constraint on the cardinality of the sequence $\{s_{i,t}\}$, or (ii) a polynomial based approach where the fact that $s_{i,t} \in \{0,1\}$ is handled via the polynomial constraint $s_{i,t}^2 = s_{i,t}$.

2.3 Time Series Comparisons and Data Driven Anomaly Detection.

The ability to quickly detect anomalies and take corrective action is a key enabler for operating in contested environments and for developing fault tolerant controllers. While anomaly detection methods have a rich history in the control community, most existing methods are model-based. Briefly, these methods make use of a model, usually linear, in a filter that is tuned to detect a particular fault; when the filter residuals (which reflect the differences between expected and observed operation) deviate sufficiently or become statistically colored, a fault is flagged. However, in many scenarios a process model is not always available and there is insufficient training data to find and validate one. In this case, there are currently fewer tools that can be brought to bear on the problem, and these methods generally require extensive tuning and/or strong priors on the statistical distribution of the data. Motivated by these difficulties, in this portion of the research we developed two data–driven fault detection approaches that do not require knowledge of the underlying process model, (i) a deterministic approach, based upon embedding the data in the positive definite manifold and using a Riemannian geometry there to compare time series, and (ii) a stochastic approach that estimates the probability that a given measurement has been generated by the same process that generated training samples drawn from the non-faulty process.

2.3.1 Comparing Time Series via Positive Definite Embeddings [11, 21]. Consider the problem of 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. 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 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. In addition, the entailed computational complexity is far from trivial. As we have shown in [11, 21] an alternative, computationally efficient method can be be obtained by first embedding the data in the positive definite manifold and then using a Riemannian manifold metric to compared the embedded data.

Briefly, given two time series $\{\mathbf{y}_t^{(1)}\}_{t=1}^{2n-1}$ and $\{\mathbf{y}_t^{(2)}\}_{t=1}^{2n-1}$, let $\mathbf{H}_{\mathbf{y}_t^{(i)}}$ denote their associated Hankel matrices. Then, it can be easily shown that, under mild observability conditions, both sequences are behaviors of the same underlying system if and only if

$$rank(\mathbf{H}_{\mathbf{y}_t^{(1)}}) = rank(\mathbf{H}_{\mathbf{y}_t^{(2)}}) = rank(\begin{bmatrix}\mathbf{H}_{\mathbf{y}_t^{(1)}} & \mathbf{H}_{\mathbf{y}_t^{(2)}}\end{bmatrix}) \tag{52}$$

While in principle this equation gives a tractable test, in practice rank is difficult to estimate in the presence of noise. Thus, rather than working directly with Hankel matrices, we embedded the data in the positive

(semi) definite manifold via the associated regularized Gram matrices defined as:

$$\mathbf{G}_\sigma = \frac{\mathbf{H}\mathbf{H}^\top}{\|\mathbf{H}\mathbf{H}^\top\|_F} + \sigma\mathbf{I} \tag{53}$$

Then, the main results of [19,20] show that, for noiseless data, two given sequences are behaviors of the same underlying system, possibly for different initial conditions, if and only if $\lim_{\sigma\to 0} JBLD(\mathbf{G}_\sigma^{(1)}, \mathbf{G}_\sigma^{(2)}) < \infty$. In the case of noisy sequences, we have shown that, if the noise is white, the JBLD between behaviors of the same system is generically much smaller than the distance between those originating from different ones. Thus, from a practical standpoint, the hypothesis that two given sequences are behaviors of the same system can be validated by simply computing the manifold distance. The advantage of this approach, in addition to its robustness to noise, is its very low computational complexity, since computing the JBLD only entails computing determinants (as opposed to computing rank that requires an SVD).

2.3.2 Anomaly Detection as a Generalized Moments Problem (GMP) [22]. The main idea here is to learn the *shape* of the non-faulty-state distribution, succinctly described by moment information and use this information to compute an upper-bound on the probability of observing a particular data sample over all the distributions that have the same moments. Briefly, in this approach all available information about the probability distribution function of the nominal process is encapsulated in a (truncated) sequence of moments of the form $m_\alpha \doteq \mathcal{E}(x^\alpha) \approx \frac{1}{N}\sum_{i=1}^N x_i^\alpha$ (recall that a probability density function is completely characterized by its moments sequence). Given a new measurement x, one can then calculate an upper bound of the probability of the observation x given the past data by solving a problem of the form:

$$\rho_{\text{mom}} = \sup_{\mu\in\mathcal{M}(\mathbb{R}^n)_+} \frac{1}{2\epsilon}\int_{x-\epsilon}^{x+\epsilon} d\mu \tag{54}$$
$$\text{s.t.} \int_{\mathbb{R}^n} x^\alpha d\mu = \gamma_j,\ j = 1,\ldots,m$$

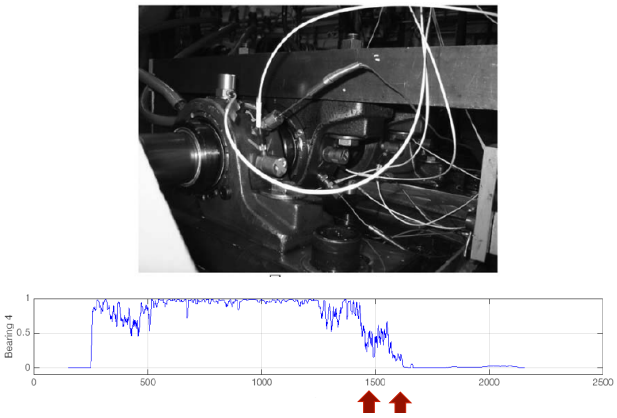


Figure 13: Data driven anomaly detection. Top: Test rig. Bottom: Probability that an observation corresponds to a non-faulty state. The fault detection threshold was crossed 5 days prior to bearing failure [22].

where $\mathcal{M}(\mathbb{R}^n)_+$ is the positive cone containing finite Borel measures μ on $\mathbb{R}^n$. As shown in [22], the problem above is a special case of the so-called generalized moments problem and can be solved via a sequence of convex relaxations. The effectiveness of this approach in detecting anomalies is illustrated in Fig. 13, where it was used to detect an incipient failure in a test rig. It is worth noting that the failure was detected substantially earlier than using existing methods.

2.4 Controller Synthesis

The ultimate goal of this research is to develop a comprehensive data-driven framework for data-driven robust decision making in contested environments. Thus, in parallel with the efforts directed towards control-oriented identification and model (in)validation, we also carried out research aimed to synthesize data-driven controllers that make optimal use of all the information available while respecting physical constraints. Interestingly, our results show that in, in many scenarios, the control synthesis problem exhibits a structure

similar to that observed in systems identification (an underlying chordal graph that is sparse). As before, exploiting this structure, together with polynomial optimization tools, leads to efficient relaxations of problems known to be, in principle, NP hard, along with optimality certificates showing that in many cases, the solution to these relaxations are indeed optimal.

2.4.1 Data Driven Robust Control [23–25]. Many practical scenarios involve designing controllers for systems where a model is not available and must be identified from a combination of experimental data and some a-priori information. Classically, this problem has been addressed by a multi-step procedure. Firstly, a suitable plant and associated worst-case identification bounds are obtained using control-oriented identification methods. If additional experimental data is available, then the uncertainty bounds can be refined through a model (in)validation step. Finally, a robust controller is synthesized that attempts to stabilize the family of plants defined by the identified nominal model and uncertainty bounds. While this multi-step approach usually works in practice, it can be overly conservative (due to the use of upper bounds in the identification and (in)validation steps). Thus, there is no guarantee that a suitable controller can be found. In addition, the entailed computational complexity is far from trivial. Motivated by these difficulties, in this portion of the research we developed a new framework for directly identifying controllers from experimental data, without explicitly identifying the plant. We have explored two different approaches: (i) imposing *superstability* and (ii) using polyhedral Lyapunov functions combined with polynomial based optimization and duality. The first approach leads to computationally tractable convex optimization problems, but can be potentially conservative, while the second is guaranteed to find a controller, if one exists, at the cost of increased computational complexity.

(i) Superstabilizing Data Driven Controllers: The SISO case [23]. Our initial approach focused on designing super-stabilizing controllers. While super-stability is a stronger requirement than Schur stability, it has the advantage of having a direct connection with time-domain, peak–to–peak performance, allowing for designing a controller with guaranteed worst case performance over the entire consistency set. The main result of this research is that the problem of synthesizing a robustly super-stabilizing controller directly from the experimental data, albeit in principle non-convex, can be recast into a convex Linear Program by exploiting results from Robust Optimization and exactly solved (see [23] for details).

Briefly, recall that a system of the form

$$y_t = -\sum_{k=1}^{n} a_k y_{t-k} + \sum_{k=0}^{n} b_k w_{t-k} \tag{55}$$

where e_t and w_t denote the output and exogenous disturbance, respectively is said to be superstable, if its characteristic polynomial $D(\lambda)$ is super stable, i.e. $\sum_1^n |a_k| < 1$. As noted above, while superstability is a stronger condition that just stability, the advantage is that it provides worst case bound on performance, that is, a superstable system satisfies $\sup_w \|y\|_\infty \leq \frac{\|b\|_1}{1-\|a\|_1}$, and this bound is tight. Further, as shown below, exploiting this bounds leads to data driven robust controllers with guaranteed performance for all plants compatible with the observed data.

Specifically, given experimental input/output data $\{u_t, y_t\}$, $\forall t = 0, \ldots, T$, where u_t and y_t denote a known input and the system output corrupted by additive process noise, respectively, e.g.

$$y_t = -\sum_{k=1}^{n_a} a_k^{(p)} y_{t-k} + \sum_{k=1}^{n_b} b_k^{(p)} (u_{t-k} + w_{t-k}) + \eta_t, \ |\eta_t| \leq \epsilon \tag{56}$$

then the consistency set is defined as the set of all plants of the form (55) that could have generated the

observed data. In the case of ℓ^∞ bounded noise, this set is a a polyhedron of the form

$$\begin{aligned}\mathcal{P}(\boldsymbol{\theta}_p) \doteq &\{\boldsymbol{\theta}_p \in \Re^{n_a+n_b} : |y_t - \mathbf{x}_t^T\boldsymbol{\theta}_p| \leq \epsilon, \forall_{t=n_a}^T\} \\ = &\{\boldsymbol{\theta}_p \in \Re^{n_a+n_b} : \mathbf{D}\boldsymbol{\theta}_p + \mathbf{q} \geq 0\}\end{aligned} \tag{57}$$

where $\mathbf{x}_t \doteq \begin{bmatrix}-y_{t-1} & \dots & -y_{t-n_a} & u_{t-1} & \dots & u_{t-n_b}\end{bmatrix}^T$ and $\boldsymbol{\theta}_p \doteq \begin{bmatrix}a_1^{(p)} & \dots & a_{n_a}^{(p)} & b_1^{(p)} & \dots & b_{n_b}^{(p)}\end{bmatrix}^T$ denotes the unknown parameters of the plant. Here, $\mathbf{D} \in \Re^{2(T+1-n_a)\times(n_a+n_b)}$ and $\mathbf{q} \in \Re^{2(T+1-n_a)}$ are known constants. In this context, the problem of interest can be precisely stated as:

Problem 10. *Given a set of experimental data $\{u_t, y_t\}_{t=0}^T$ collected from a plant P of the form* (56) *with known order n_a and n_b, and a bound ϵ on ℓ_∞ norm of the noise, design a controller C (determined by unknown $\boldsymbol{\theta}_c = [a_1^{(c)}, \cdots, a_{n_c}^{(c)}, b_0^{(c)}, \cdots, b_{n_c}^{(c)}]^T$) with the transfer function*

$$C(\lambda) = \frac{N_c(\lambda)}{D_c(\lambda)} = \frac{\sum_{k=0}^{n_c} b_k^{(c)}\lambda^k}{1 + \sum_{k=1}^{n_c} a_k^{(c)}\lambda^k} \tag{58}$$

such that the resulting closed-loop system is superstable for any plant in the consistency set $\mathcal{P}(\boldsymbol{\theta}_p)$, and the worst case ℓ^∞ induced norm of the closed–loop system is minimized.

To reformulate Problem 10 as an optimization problem, begin by writing explicitly the closed-loop transfer function of the system as:

$$T(\lambda) = \frac{P(\lambda)}{1 + P(\lambda)C(\lambda)} = \frac{N_pD_c}{D_pD_c + N_pN_c} \doteq \frac{\sum_{k=1}^{n} b_k\lambda^k}{1 + \sum_{k=1}^{n} a_k\lambda^k} \tag{59}$$

where the unknown coefficients $\boldsymbol{\theta} = \begin{bmatrix}\mathbf{a}^T & \mathbf{b}^T\end{bmatrix}^T \in \Re^{2n+1}$ of $T(\lambda)$ are bilinear functions of the parameters of the plant and controller, $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_c$ respectively, of the form:

$$\theta_i = (\mathbf{Q}_i\boldsymbol{\theta}_p + \mathbf{f}_i^{(c)})^T\boldsymbol{\theta}_c + \mathbf{f}_i^{(p)T}\boldsymbol{\theta}_p + r_i, \forall_{i=1}^{2n+1} \tag{60}$$

where $\mathbf{Q}_i \in \Re^{(2n_c+1)\times(n_a+n_b)}$, $\mathbf{f}_i^{(c)} \in \Re^{2n_c+1}$, $\mathbf{f}_i^{(p)} \in \Re^{n_a+n_b}$, and $r_i \in \Re$ are known constants. Using this notation, Problem 10 can be reformulated as a constrained optimization problem of the form

$$\tilde{\mu}^* = \min_{\boldsymbol{\theta}_c, \mu>0} \mu \tag{61a}$$

$$\text{s.t. } \mu||\mathbf{a}||_1 + ||\mathbf{b}||_1 \leq \mu, \forall\boldsymbol{\theta}_p \in \mathcal{P}(\boldsymbol{\theta}_p) \tag{61b}$$

In principle, problem (61) is nonconvex due to the bilinear terms. However, as we have shown in [23], surprisingly, application of the extended Farkas Lemma allows for reducing it to a Linear Program of the form:

$$\begin{aligned}&\min \mu \text{ subject to:} \\ \mathbf{q}^T\mathbf{z}_i + \bar{\mathbf{f}}_i^{(c)T}\boldsymbol{\theta}_c + \bar{r}_i \leq 0 \\ -\mathbf{D}^T\mathbf{z}_i = \bar{\mathbf{f}}_i^{(p)} + \bar{\mathbf{Q}}_i^T\boldsymbol{\theta}_c \\ \mathbf{z}_i \geq \mathbf{0} \\ \forall i = 1, \cdots, 2^{2n+1},\end{aligned} \tag{62}$$

Further, as we have shown in [23] this relaxation is exact, that is, there is no duality gap.

(ii) Extension to Switched MIMO Systems and Switched Difference Inclusions [24, 25]. As indicated in section 2.1.4, switched systems have been the object of renewed interest in the past decade, since they allow for handling scenarios where the dynamics change, for instance as the operating conditions change. In particular, a class of problem of interest to this research is switched difference inclusions, where, due to uncertainty, at any given time, the system is only known to belong to a given set, and the goal is to render the origin an asymptotically equilibrium point, regardless of the switching sequence. An example of scenarios where this approach is relevant is fault tolerant control, where the controller needs to drive the system to a safe state, in the presence of faults whose dynamics were previously unknown. Specifically, we sought to solve the following problem:

Problem 11. *Consider the setup shown in Figure 14 and assume that the system under consideration consists of s LTI subsystems, each described by a model of the form:*

$$\mathbf{x}_{k+1}^i \quad = \quad \mathbf{A}_i\mathbf{x}_k^i + \mathbf{B}_i\mathbf{u}_k + \mathbf{w}_k,\ i \in \{1,\dots,s\} \tag{63}$$

Given labeled experimental data $\{u_k, x_k^i, x_{k+1}^i\}_{k=0}^T$, $i = 1,\dots,s$, *and a bound* ϵ *on* ℓ_∞ *norm of the noise, find a switched state feedback gain* $\mathbf{F}_i$ *such that the resulting closed loop matrix* $\mathbf{A}_{\sigma_t} + \mathbf{B}_{\sigma_t}\mathbf{F}_{\sigma_t}$ *is asymptotically stable for any switching sequence* $\sigma_t \in \{1,\dots,s\}$, *for all pairs* $(\mathbf{A}_i, \mathbf{B}_i)$ *in the consistency set* $\mathcal{P}$ *defined as the set of all plants of the form* (63) *that could have generated the observed data.*

We solved this problem by exploiting the fact that the origin is an asymptotically stable equilibrium point of (63) under arbitrary switching if and only if there exists a full column rank matrix [2] $\mathbf{V}$ and n_s matrices $\mathbf{H}_i$, $\|\mathbf{H}_i\|_\infty < 1$ such that

$$\mathbf{V}\mathbf{A}_i = \mathbf{H}_i\mathbf{V},\ i = 1,\dots,n_s \tag{64}$$

Next, note that since (63) is linear in $\mathbf{A}, \mathbf{B}$ and the noise is bounded in the ℓ^∞ sense, the consistency set $\mathcal{P} = \cup\mathcal{P}^{(i)}$, where each of the $\mathcal{P}^{(i)}$ is a polytope of the form:

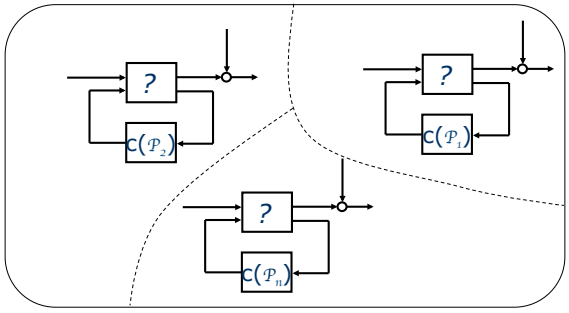


Figure 14: Setup for Switched Data Driven Control Synthesis.

$$\begin{aligned}\mathcal{P}^{(i)} \doteq \{\mathbf{a}_j^{(i)}, \mathbf{b}_j^{(i)} : \| &\left(\mathbf{x}_{t_1^{(i)}}^T \otimes \mathbf{I}\right)\mathbf{a}_j^{(i)} + \\ &\left(\mathbf{u}_{t_1^{(i)}}^T \otimes \mathbf{I}\right)\mathbf{b}_j^{(i)} - \mathbf{x}_{t_1^{(i)}+1}\|_\infty \le \epsilon\}\end{aligned} \tag{65}$$

where $(\mathbf{A}_j^{(i)}, \mathbf{B}_j^{(i)})$ denotes a generic pair in $\mathcal{P}^{(i)}$, $\mathbf{a}_j^{(i)} \doteq \mathbf{vec}(\mathbf{A}_j^{(i)})$, $\mathbf{b}_j^{(i)} \doteq \mathbf{vec}(\mathbf{B}_j^{(i)})$ and where $t_k^{(i)}$, $k = 1,\dots,m_i$ denotes the times at which the i^{th} system is active. In this context, Problem 11 is equivalent to the following robust optimization:

Problem 12. *Find a full rank matrix* $\mathbf{V} \in \mathbb{R}^{n\times n}$, *a switched feedback gain* $\mathbf{F}_i$ *and matrices* $\mathbf{H}_j^{(i)}$ *such that*

$$\mathbf{V}(\mathbf{A_j^{(i)}} + \mathbf{B_j^{(i)}}\mathbf{F_i}) = \mathbf{H}_j^{(i)}\mathbf{V} \text{ and } \|\mathbf{H}_j^{(i)}\|_\infty \le d < 1 \tag{66}$$

for all pairs $(\mathbf{A}_j^{(i)}, \mathbf{B}_j^{(i)}) \in \mathcal{P}^{(i)}$, $i = 1,\dots,n_s$.

[2] In this case the function $\mathcal{V}(\mathbf{x}) \doteq \|\mathbf{V}\mathbf{x}\|_\infty$ is a polyhedral Lyapunov function for (**??**).

The main result of this portion of the research showed that, through the use of duality Problem 12 can be recast into a polynomial optimization form. Specifically, denote by $t_k^{(i)}$, $k = 1, \ldots, m_i$ the time instants where the i^{th} sub-system is active. Let:

$$\mathcal{X}^{(i)} \doteq \begin{bmatrix} \mathbf{x}_{t_1^{(i)}}^T \otimes \mathbf{I} \\ \vdots \\ \mathbf{x}_{t_{m_i}^{(i)}}^T \otimes \mathbf{I} \end{bmatrix}, \mathcal{U}^{(i)} \doteq \begin{bmatrix} \mathbf{u}_{t_1^{(i)}}^T \otimes \mathbf{I} \\ \vdots \\ \mathbf{u}_{t_{m_i}^{(i)}}^T \otimes \mathbf{I} \end{bmatrix}, \boldsymbol{\xi}^{(i)} \doteq \begin{bmatrix} \mathbf{x}_{t_k^{(i)}+1} \\ \vdots \\ \mathbf{x}_{t_{m_i}^{(i)}+1} \end{bmatrix}$$

We have shown (see [25] for details) that, given a (full rank) matrix $\mathbf{V} \in \mathbb{R}^{n \times n}$, a matrix $\mathbf{S} \in \mathbb{R}^{r \times n^2}$ and a non-negative vector $\boldsymbol{\lambda} \in \mathbb{R}^r$ there exist switched feedback gains $\mathbf{F}_i$ and matrices $\mathbf{H}_j^{(i)}$ such that

$$\mathbf{V}(\mathbf{A}_\mathbf{j}^{(\mathbf{i})} + \mathbf{B}_\mathbf{j}^{(\mathbf{i})}\mathbf{F}_\mathbf{i}) = \mathbf{H}_j^{(i)}\mathbf{V} \text{ and } \mathbf{S}\text{vec}(\mathbf{H}_j^{(i)}) \leq \boldsymbol{\lambda} \tag{67}$$

for all pairs $(\mathbf{A}_j^{(i)}, \mathbf{B}_j^{(i)}) \in \mathcal{P}^{(i)},\ i = 1, \ldots, n_s$ if and only if there exist n_s matrices $\mathbf{Y}^{(i)} \in \mathbb{R}^{r \times 2nm_i},\ \mathbf{Y}^{(i)} \geq 0$ and $\mathcal{F}_i \in \mathbb{R}^{m \times n}$ such that

$$\begin{aligned} &\mathbf{Y}^{(i)} \begin{bmatrix} \mathcal{X}^{(i)} & \mathcal{U}^{(i)} \\ -\mathcal{X}^{(i)} & -\mathcal{U}^{(i)} \end{bmatrix} = \begin{bmatrix} (\mathbf{S}(\mathbf{V}^{-T} \otimes \mathbf{V}) & \mathbf{S}(\mathcal{F}_i^T \otimes \mathbf{V}) \end{bmatrix} \\ &\mathbf{Y}^{(i)} \begin{bmatrix} \boldsymbol{\xi}^{(i)} + \epsilon \mathbf{1} \\ -\boldsymbol{\xi}^{(i)} + \epsilon \mathbf{1} \end{bmatrix} \leq \boldsymbol{\lambda} \end{aligned} \tag{68}$$

Exploiting this result allowed for recasting Problem 12 into the following polynomial feasibility problem: s equivalent to the following polynomial feasibility problem: Find n_s matrices $\mathbf{Y}_i \in \mathbb{R}^{n \times 2nm_i} \geq 0, \mathcal{F}_i \in \mathbb{R}^{n \times m}$, a full rank matrix $\mathbf{V} \in \mathbb{R}^{n \times n}$ and a non-negative vector $\boldsymbol{\lambda}$ with elements $\lambda_i \leq d < 1$ such that (68) holds for all matrices $\mathbf{S} \in \mathbb{R}^{n \times n^2}$ of the form:

$$\mathbf{S} = \begin{bmatrix} \text{diag}(\mathbf{s}_1) & \text{diag}(\mathbf{s}_2) & \ldots \text{diag}(\mathbf{s}_n) \end{bmatrix} \tag{69}$$

where $\mathbf{s}_i \in \mathbb{R}^n$ is a vector with elements $s_{i,j} = \pm 1$. In turn, this problem can be solved using Algorithm 4 listed below, obtained using polynomial optimization techniques:

2.4.2 Synthesis of controllers subject to information flow constraints [26–28] The past decade has witnessed an explosive growth in sensing and actuation capabilities, rendering feasible distributed control applications involving a very large number of networked sensors and actuators. However, fully exploiting the capabilities of these distributed systems requires developing control synthesis methods that take into account structural interconnection constraints. These arise for example from constraints on the number of interconnections that can be handled at each node or the need to respect physical constraints. It is well known that this problem is generically NP-hard, unless the plant satisfies the Quadratic Invariance property. Unfortunately, this property does not hold for generic information structures. Motivated by this difficulty, in this portion of the research we developed efficient, tractable convex relaxations for synthesizing information flow constrained controllers. Specifically, we addressed the following problem:

Problem 13. *Given system of the form*

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}_\mathbf{v}\mathbf{v}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}_\mathbf{v}\mathbf{v}(t) \end{aligned} \tag{70}$$

where $\mathbf{A} \in \mathbb{R}^{r \times r}$, $\mathbf{B} \in \mathbb{R}^{r \times p}$, $\mathbf{B}_\mathbf{v} \in \mathbb{R}^{r \times e}$, $\mathbf{C} \in \mathbb{R}^{q \times r}$ *and* $\mathbf{D}_\mathbf{v} \in \mathbb{R}^{q \times e}$ *are known matrices, and where* $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ *and* $\mathbf{v} \in \mathbb{R}^e$ *represent the control input, the measured output and disturbances, design a stabilizing controller subject to constraints on the information exchange between sensors and actuators,*

$$\mathbf{u}(t) = \mathbf{K}\mathbf{y}(t) \text{ and } \mathbf{K} \in \mathbf{S} \tag{71}$$

where S *is a given sparsity structure.*

Algorithm 4 Reweighted $\|.\|_*$ based DDC design

Initialize: $iter = 0, \mathbf{W}^{(0)} = I$, $d < 1$

repeat

Solve

$$\begin{aligned}
&\min_m \quad \text{Trace}(\mathbf{W}^{(iter)}\mathbf{M}) \\
&\text{subject to:} \\
&\boldsymbol{\lambda} \geq 0 \\
&\mathbf{M}(m) \succeq 0 \\
&\mathbf{M}(1,1) = 1 \\
&\mathbf{VZ} = \mathbf{I} \\
&\mathbf{mat}(\boldsymbol{\lambda})\mathbf{1} \leq d\mathbf{1} \\
&\text{and, for all } i = 1, \ldots, n_s \\
&\mathbf{Y}^{(i)} \geq 0 \\
&\mathbf{Y}^{(i)} \begin{bmatrix} \mathcal{X}^{(i)} & \mathcal{U}^{(i)} \\ -\mathcal{X}^{(i)} & -\mathcal{U}^{(i)} \end{bmatrix} = \begin{bmatrix} \mathcal{K} & \mathcal{N}^{(i)} \\ -\mathcal{K} & -\mathcal{N}^{(i)} \end{bmatrix} \\
&\mathbf{Y}^{(i)} \begin{bmatrix} \boldsymbol{\xi}^{(i)} + \epsilon\mathbf{1} \\ -\boldsymbol{\xi}^{(i)} + \epsilon\mathbf{1} \end{bmatrix} \leq \boldsymbol{\lambda}
\end{aligned}$$

where $\mathcal{K} = \mathbf{Z}^T \otimes \mathbf{V}, \mathcal{N}^{(i)} = \mathcal{F}_i^T \otimes \mathbf{V}$

$\mathbf{M}$ represents the moment matrix given by:

$\mathbf{M} = [1, \mathbf{vec}(\mathbf{V})^T, \mathbf{vec}(\mathbf{Z})^T, \mathbf{vec}(\mathcal{F}_i)^T]^T \times$
$[1, \mathbf{vec}(\mathbf{V})^T, \mathbf{vec}(\mathbf{Z})^T, \mathbf{vec}(\mathcal{F}_i)^T]$

Update

$$\begin{aligned}
&\mathbf{W}^{(iter+1)} = (\mathbf{M}^{(iter)} + \sigma_2(\mathbf{M}^{(iter)})I)^{-1} \\
&iter = iter + 1
\end{aligned}$$

until $rank\{\mathbf{M}\} = 1$.

We have developed two different approaches to solve this problem: the first method is based upon the idea of recasting the problem into a sparse filtering form while the second exploits a combination of polyhedral Lyapunov function and polynomial optimization ideas.

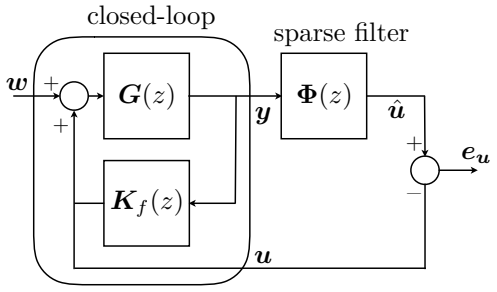


Figure 15: Sparse controller design as a sparse filtering problem.

(i) Sparse controller design as a filtering problem [26]. The main idea here, illustrated in Fig. 15, is to first design a non-sparse controller K_f and then find the best sparse approximation to the optimal control action (rather than the controller), $u = \Phi(z)y$. As shown in [26], finding the optimal sparse "filter" $\Phi(z)$ can be recast into the following convex form:

$$J = \min_{\mathbf{\Phi}(z)\in\mathcal{RH}_\infty\cap\mathcal{S}} \|\mathbf{T_{uw}}(\mathbf{K}_f) - \mathbf{\Phi}\mathbf{T_{yw}}(\mathbf{K}_f)\|_{\mathcal{H}_\infty} \tag{72}$$

where $\mathbf{T_{uw}}(\mathbf{K}_f) = \mathbf{K}_f(\mathbf{I} - \mathbf{G}\mathbf{K}_f)^{-1}\mathbf{G}$ and $\mathbf{T_{yw}}(\mathbf{K}_f) = (\mathbf{I} - \mathbf{G}\mathbf{K}_f)^{-1}\mathbf{G}$ denote the closed-loop transfer functions from the disturbance input $\mathbf{w}$ to the control action $\mathbf{u}$ and measured output $\mathbf{y}$, respectively.

(ii) Sparse controller design via polynomial optimization [27, 28]. The approach described above works well when it is desired to approximate a stable controller K_f. However, in many cases, the desired controller is not open-loop stable (for instance if the plant is not strongly stabilizable). To address this issue, we have also developed an approach that can handle generic systems, at the price of higher computational complexity. The starting point is to use a polyhedral Lyapunov function to parameterize the set of all stabilizing controllers that have the desired structure. Consider first the case of static controllers. As shown in [27,28], a static output-feedback controller $\mathbf{K}$ stabilizes the system (70) if and only if there exist matrices $\mathbf{P}$ and $\mathbf{H}$ and a scalar δ, such that

$$\begin{bmatrix}\mathbf{I}\\ \mathbf{P}\end{bmatrix}(\mathbf{A}+\mathbf{BKC}) = \mathbf{H}\begin{bmatrix}\mathbf{I}\\ \mathbf{P}\end{bmatrix}$$
$$\|\mathbf{H}\|_1 \le \delta,\ 0 \le \delta < 1 \tag{73}$$

Further, a given sparse structure $\mathbf{S}$ can be directly incorporated into the parameterization (73) by simply enforcing $\mathbf{K}(i,j) = 0,\ \forall (i,j)$ such that $\mathbf{S}(i,j) = 0$. Since all the conditions above only involve polynomials, a suitable $\mathbf{K}$ can be found by using moments (or SoS) techniques to reduce the problem to a semi-definite program. Next, we briefly outline how to extend these ideas to dynamic controllers. Consider a controller of the form

$$\begin{aligned}\mathbf{x}_u(t+1) &= \mathbf{A}_u\mathbf{x}_u(t) + \mathbf{B}_u\mathbf{y}(t)\\ \mathbf{u}(t) &= \mathbf{C}_u\mathbf{x}_u(t) + \mathbf{D}_u\mathbf{y}(t)\end{aligned} \tag{74}$$

with $\mathbf{A}_u \in \mathbb{R}^{r_u\times r_u}$. The corresponding closed loop dynamics are given by:

$$\begin{bmatrix}\mathbf{A}+\mathbf{B}\mathbf{D}_u\mathbf{C} & \mathbf{B}\mathbf{C}_u\\ \mathbf{B}_u\mathbf{C} & \mathbf{A}_u\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \\ & \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{B} & \\ & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{D}_u & \mathbf{C}_u\\ \mathbf{B}_u & \mathbf{A}_u\end{bmatrix}\begin{bmatrix}\mathbf{C} & \\ & \mathbf{I}\end{bmatrix} \tag{75}$$

Since this is similar to the formulation $\mathbf{A}+\mathbf{BKC}$ derived for the static control case, the same tools used in that case be directly applied here. Finally, we note that the ideas outlined above can be combined with tools from viability theory to synthesize sparse controllers subject to hard control constraints. Details are provided in [27].

3 Applications

Although the main focus of this effort is basic research, in order to illustrate the advantages of the proposed framework, we have applied it to solve several non-trivial practical problems.

3.1 Applications of the parsimonious models identification framework

Next, we illustrate the potential applications of the parsimonious model identification outlined in section 2.1.2 to three different domains (i) damage mitigating control, (ii) tracking in contested environments, and (iii) multi-sensor data fusion.

3.1.1 Identification of a lightly damped system for damage mitigation control [1, 3, 4].

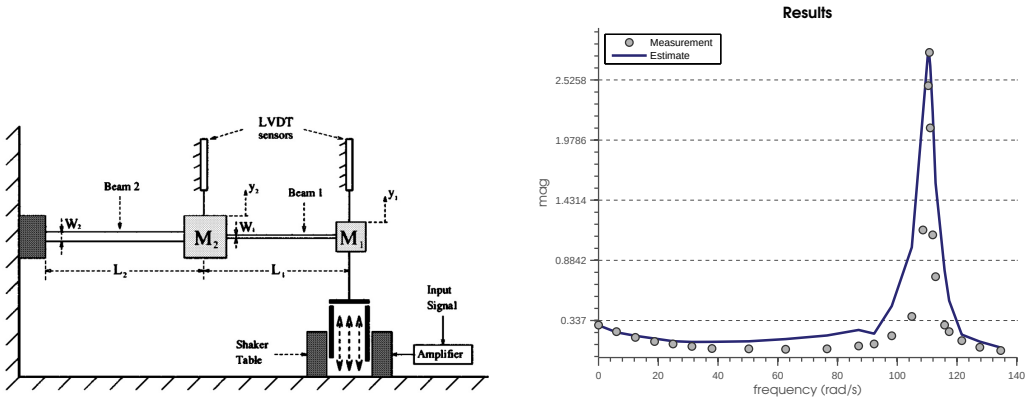


Figure 16: Left: A lightly damped system used for damage mitigation control testing. Right: Experimental data vs identified model fit.

Damage mitigation control seeks to extend the life of compromised structures by switching from a nominal controller to a less aggressive one that trades off performance versus stress with the goal of allowing for a damaged vehicle to return to a safe base or a damaged structure to remain in place until it can be replaced. Central to this concept is the ability to, once damage occurs, quickly identify a new model of the process and synthesize a new controller. Since both of these must be done in real time, computational complexity (which as indicated before is dominated by the order of the model) must be kept as low as possibly, compatible with the requirement of fidelity of the model to the new data. Consider the problem of identifying the very lightly damped two-degrees of freedom structure shown in Fig. 16, used as a benchmark for damage mitigating control, from the experimental data shown there. Note the very large resonance peak that renders the identification problem non-trivial. Using the proposed atomic norm approach leads to a 6^{th} order system that fits well the data, in approximately 2 seconds. For comparison, an ADMM based approach, while producing similar results, requires approximately 2500 seconds [7].

3.1.2 Multi radar target tracking in contested environments [29, 30]. In any adversarial environment, it is important to keep a good view of the air space, since this information can be used as early warning for headquarters, cities, and for deploying countermeasures. Most of the time, this air picture is generated by surveillance radars. Since having a good and long range view creates advantage for the defending side, adversaries usually try to interfere with radars and hence with the estimated air picture. One common practice is the use of jamming air vehicles. In a jammed environment, surveillance radars might lose the range

information and only azimuth (direction) or bearing information can be measured.

Such an ambiguity in range makes early deployment of countermeasures very hard if not impossible. On the other hand, it is possible to localize target jammers if there is more than one radar. In this case, it is well known in the tracking literature that one can use triangulation methods to track jammers. However, triangulation may produce extraneous intersections which are called "ghosts" (see Fig 17). Unfortunately, these ghosts move and behave like jammers, and so the radars can not decide where are the correct targets. The goal of this portion of the research was to develop a computationally tractable method to separate the real jamming vehicles from the "ghosts". The key observation is that "ghost" trajectories are a function of trajectories of several "real" targets and, hence, we expect them to have a substantially higher level of complexity. Further, the number of "ghosts" n_g is generically known from the geometry of the scenario. Thus, "ghosts" can be distinguished from actual targets by using the atomic norm identification framework outlined in section 2.1.1 to identify a parsimonious model for each trajectory and selecting the n_g models that have the highest order [29].

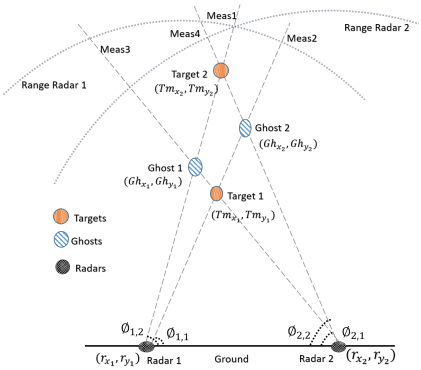


Figure 17: Multi-target tracking using bearing information only. Targets can be distinguished from ghost based on the complexity of their motion.

3.1.3 Solving "Temporal Puzzles" (multi-sensor data fusion). Most of the problems of interest in this research are based on having an ordered temporal sequence of data $y(t)$. Once this sequence is available, the methods outlined in section 2 can be used to identify a model, detect anomalies and synthesize controllers. However, in many practical scenarios, only "snapshots" y_i of the temporal sequence are available, with only a rough time stamp. Examples include trying to reconstruct an activity or track a target from pictures taken by different sensors (such as individual cell-phones). In these cases, a pre-requisite to identification is to reconstruct the sequence $y(t)$ from the un-ordered samples y_i. Clearly, as stated this problem is ill posed, since given any ordering of n data points, one can always find an $n-1$ order system that interpolates them. On the other hand, in the absence of additional priors, the ordering that leads to the lowest order interpolant is usually the correct one (see Fig. 18). Combining this observation with the atomic norm framework outlined in section 2.1.2 leads to the following regularized problem [31]:

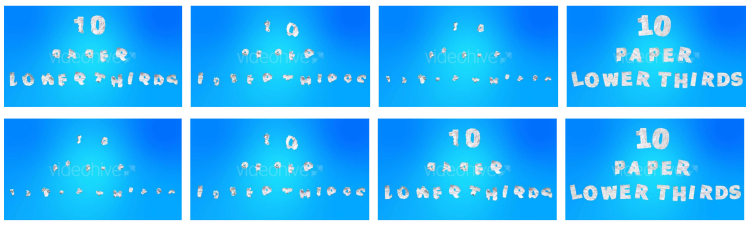


Figure 18: Solving temporal puzzles via atomic norm minimization. Top: unsorted data. Bottom: sorted data. The correct ordering corresponds to the lowest order interpolating model.

$$\begin{array}{ll} \text{minimize}_{\mathbf{c}} & \|\mathbf{c}\|_{\ell_1} \\ \text{subject to} & \mathbf{y}_s = \mathbf{D_a c} \\ & \|\mathbf{P}\mathbf{y}_s - \mathbf{y}_u\|_\infty \leq \eta_{max}, \;\; \mathbf{P} \in \mathcal{P} \end{array} \tag{76}$$

Here $\mathbf{y}_s$ and $\mathbf{y}_u$ denote vectors containing the sorted and unsorted data, $\mathcal{P}$ is the set of permutation matrices of appropriate dimension, η_{max} is a bound on the size of the noise, and $\mathbf{D_a}$ is a matrix whose columns are the impulse response of atoms in a suitable dictionary $\mathcal{A}$, (for instance systems having poles inside an origin

center ring of radius r). The resulting problem is a MILP that can be efficiently solved using off the shelf solvers such as Gurobi that can handle very large problems. Note in passing that any *a priori* partial ordering information available can be easily incorporated by introducing an auxiliary vector, $\mathbf{l} = [1,\ 2,\ \ldots,\ T]^T$. If the i^{th} data point is known to precede the j^{th} input data point for a given sorting problem, this can be enforced by adding the constraint $\mathbf{P_i}^T\mathbf{l} < \mathbf{P_j}^T\mathbf{l}$ to (76), where $\mathbf{P_i}$ denotes the i^{th} column of $\mathbf{P}$.

3.2 Finding causally correlated activities in video sequences [10]

The goal of this portion of the research was to explore the use of dynamical graphical models to unveil causal relationships from time series generated by different agents. In particular, we analyzed video sequences from the UT Human Interaction Data Set. The specific time series used in this example are the trajectories of each individual's head, normalized to lie in the interval $[-1, 1]$. Figure 19 shows the result of applying Algorithm 2 (section 2.1.3), modified to take into account the existence of derivative-sparse exogenous inputs, to two sample sequences. As shown there, this algorithm successfully identified the interactions between agents in both sequences. In addition, the super-atomic norm based approach was, depending on the examples, 3 to 5 times faster than using an ADMM based algorithm [10].

Figure 19: Sample Frames of the UT Sequences 6 and 16 showing the causally interacting groups identified using the approach outlined in Section 2.1.3 [10].

3.3 Activity analysis from noisy video data [5]

The goal of this application is to segment a video containing multiple activities into sub-activities, each characterized by an affine model. As outlined in section 3.5 below, identifying these models is a pre-requisite to recognize contextually abnormal activity. The experimental data used in this application, illustrated in Figure 20, consist of 55 frames extracted from a video sequence of a person walking, bending and resuming walking. To simulate a realistic scenario, several frames were corrupted

Figure 20: Sample frames for the activity segmentation application.

with large amounts of noise, consistent with a scenario where the data is corrupted by interference. In order to recast the segmentation problem into an identification form, the position of the center of mass of the person in each frame was modeled as the output of a switched affine system consisting of 2 first order submodels, and the system was identified using the algorithm outlined in Section 2.1 4. As shown in Fig. 21, this approach successfully segmented the sequence in the presence of outliers.

3.4 Anomaly Detection

Next, we discuss the application of the ideas presented in this report to the problem of extracting actionable information from large data sets. Framing this problem using concepts from dynamical systems, allows us to exploit the tractable relaxations discussed above. This approach leads to scalable, computationally tractable algorithms, which can help making critical decisions based on dynamic information that is very sparsely encoded in the available data streams in real time. To this effect, we considered the observed data as the output of a switched dynamical system, where jumps between systems indicate events that can be characterized by the parameters of the corresponding subsystems.

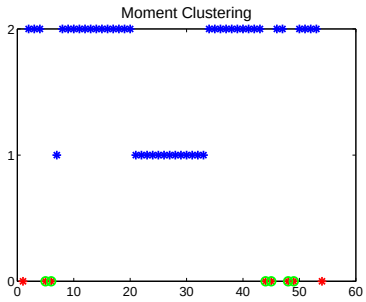


Figure 21: Activity segmentation as a Switched-ARX identification problem with outliers. (red stars and green circles denote the detected and true outliers, respectively).

Application 1: abnormal activity detection from video sequences [5, 7]. This problem, which arises in surveillance systems of large public spaces. is very challenging, since the video data usually contains many different activities. Thus, most machine learning based techniques must parse the data into sub-activities before they can detect an anomaly. Furthermore, the parsing can be very difficult if the individual segments are very short, spanning only a few frames. However, explicit parsing can be avoided by formulating the problem as a model (in)validation one: first, identifying models corresponding to "safe activities" and then detecting anomalies by using the approach outlined in Section 2.2 to invalidate the observed data against the set of trajectories that could have been generated by switching among these safe activities. This approach is illustrated by an example with three safe activities $\{$*waiting, walking, running*$\}$, described by the models (see [7]):[3]

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.4747 & 0.0628 \\ -0.3424 & 1.2250 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 0.5230 & -0.1144 \\ 0.3574 & -0.2513 \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} \quad \text{(walking)}$$

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} \quad \text{(waiting)}$$

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.6058 & 0.0003 \\ 0.2597 & 0.8589 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 0.3608 & 0.1853 \\ -0.2381 & 0.1006 \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} \quad \text{(running)}$$

[3] These models were identified by considering the trajectories of the centroid of the person and using LP to find the coefficients that minimized the peak value of the fitting error. Note that both *walking* and *running* have two poles at 1, corresponding to constant velocity motion. However, the remaining two poles for the *running* model are complex conjugate, corresponding to oscillatory motion, while those in the *walking* model are real.

where (x_t, y_t) are the coordinates of the centroid of the actor, and where transitions from *waiting* to *running* are not allowed, as shown in Fig. 22. The proposed approach successfully flagged the sequence in Fig. 23 as "anomalous", even though it all the sub-activities are safe activities, because it exhibits a forbidden transition.

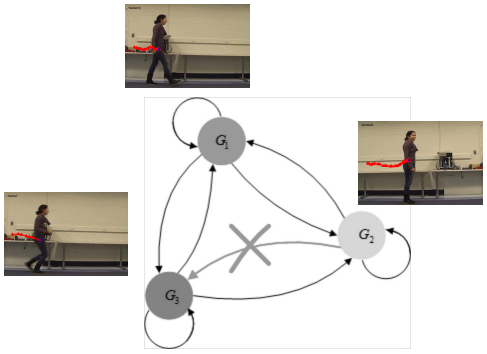


Figure 22: A structurally constrained transition graph. One-step transitions from *waiting* to *running* are not allowed.

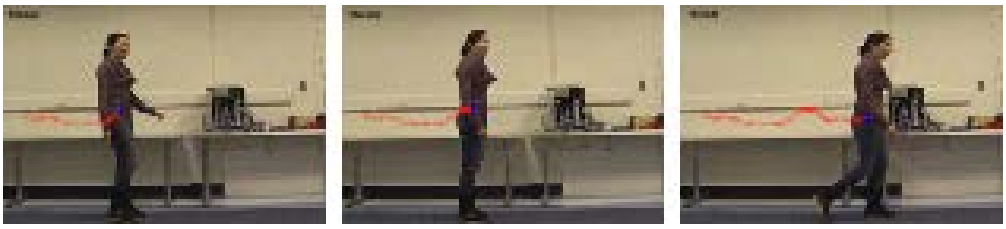


Figure 23: Anomalous behavior detection as a switched (in)validation problem. The video sequence is flagged as abnormal, even though it consists of "normal" activities (*walking,waiting,running*), since it contains an anomalous transition (*waiting* $\rightarrow$ *running*).

Application 2: fault detection in refrigerated containers [32,33]. Refrigeration containers, are similar to conventional freight containers, but are fitted with insulated walls and a refrigeration system. The advantage over conventional containers, is that the atmosphere inside is controllable which helps ensure the quality and integrity of the product throughout the supply chain. However, this also introduces a potential failure point, since a malfunction in the refrigeration mechanism would cause the contents to spoil, potentially leading to catastrophic consequences in cases where this cargo is critically needed (for instance medicines or food needed in the aftermath of a hazzard). Therefore, it is of utmost importance, to be able to detect errors, and ensure the integrity of the refrigeration system throughout the entire journey. In this portion of the research we have shown (see Fig. 24) that the anomaly detection methods described in sections 2.1.5 and 2.3.2 can successfully detect the early onset of failures (allowing for implementing corrective actions), separating these anomalies from the normal fluctuations in sensor readings.

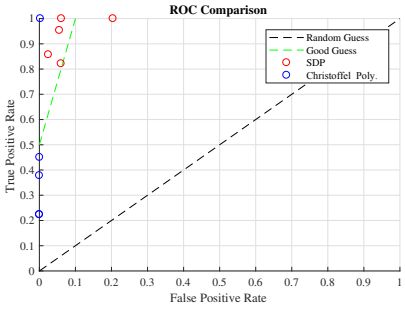


Figure 24: Operating curves for detecting anomalies in a refrigeration container. These curves were obtained using actual container data

.

3.5 Data Driven Control of a Rotary Wing UAV [34]

The area of unmanned aerial vehicles (UAVs) has seen rapid growth, mainly because of the ability of UAVs to effectively carry out a wide range of missions such as search and rescue, surveillance, power plants inspection, and battle damage assessment, without risking human life. However, successfully carrying out these missions require the ability to autonomously move in unknown environments subject to dynamic and physical constraints. Accomplishing this task requires addressing the following two issues: (i) generation of a *safe* trajectory capable in principle to reach the target by avoiding obstacle occurrences along the path, and (ii) finding a control action whose aim is to drive "as close as possible" the UAV to the planned path under the satisfaction of the constraints arising from the specific vehicle dynamics. In this portion of the research we developed a controller, based on Command Governor ideas, capable of avoiding obstacles whose presence is a priori-unknown and must be detected on-line, while respecting the physical constraints imposed by the UAV platform. The proposed architecture, shown in Fig. 25, consists of a **Planner unit**, tasked with computing a set of waypoints to be used as references for the UAV until the target is reached, a **Control Module (SCG-OA)** and a **CG-equipped control System**. The combined action of these last two modules guarantees that waypoints are reached in a finite time, without constraint violations or obstacle collisions. The **Control Module (SCG-OA)** unit is tasked with updating the current waypoint or (via switching arguments) and a convex approximation of the obstacle-free region. This convex approximation is obtained by replacing the actual non-convex obstacle free region by a sequence of properly overlapped convex approximations, computed on-line.

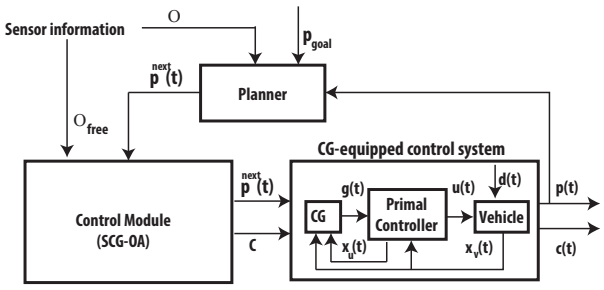


Figure 25: Control Architecture

The **CG** module consists of two nested control loops (see Fig. 25). The internal loop is designed via a generic linear control method, without taking into account the prescribed constraints, and allows the designer to specify relevant system properties for small-signal regimes, e.g. stability, disturbance rejection. The outer loop consists of the CG unit which, whenever necessary, is in charge of modifying the reference to be applied to the closed-loop system so as to avoid constraint violation. The complete computational complexity of the controller is shown in Table 1. Here **CG re-design** indicates a situation where the command-governor must be re-designed in order to maintain viability of the overlapping constraints (Fig 26), **CP** denotes the convexification procedure used to approximate the obstacle free region with a sequence of overlapping convex sets, and **SCG-OA** refers to the obstacle-avoidance trajectory planning.

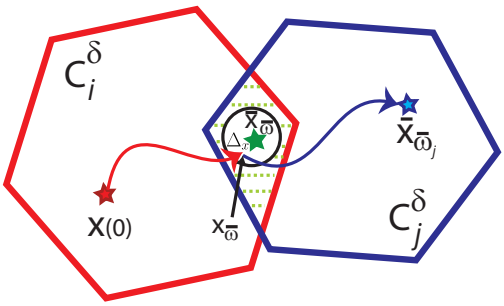


Figure 26: Viability retention under constraints change

As shown in the Table, as long as the number of obstacles is not too large, this computational complexity remains moderate, allowing for implementation on platforms with limited computational resources. The effectiveness of this approach was validated in our lab using a Qball-X4 quadrotor (see Fig. 27 and [34] for details).

Procedure	Best-Case	Worst-Case
CG re-design	0	k_0 LP opts.
CP	$\mathcal{O}(l \log l)$	$\mathcal{O}(l^2 \log l)$
SCG-OA	2 QP opts.	2 QP opts. + k_0 LP opts. + $\mathcal{O}(l^2 \log l)$

Table 1: Computational Cost: **CG** re-computation, ***CP*** procedure, and ***SCG-OA*** algorithm. Here l denotes the number of obstacles.

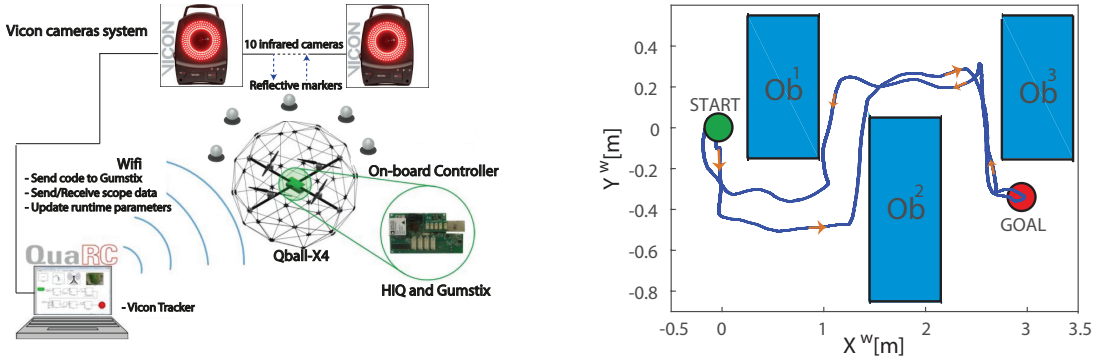


Figure 27: CG based control of an UAV. Left: experimental set-up. Right: a sample trajectory.

4 Personnel Supported, Keynote Presentations and Outreach:

4.1 Personnel Supported

Mario Sznaier	Dennis Picard Trustee Professor of ECE, Northeastern University (PI)
Octavia Camps	Professor of ECE, Northeastern University (Co-PI)
Yin Wang	Ph.D. student, graduation date, May 2016
Yongfang Cheng,	Ph.D. student, graduation date, August 2016
Jose Lopez	Ph.D. student, graduation date, December 2016
Xikang Zhang	Ph.D. student, graduation date, May 2018
Bengisu Ozbay	Ph.D. student, expected graduation date, May 2020
Tianyu Dai	Ph.D. student, expected graduation data, May 2021

4.2 Keynote and Plenary Talks

- Mario Sznaier delivered plenary talks at the 2015 Geometric and Numerical Foundations of Movement; 2016 Methods and Algorithms for Control (MAC); 2017 Benelux Meeting on Systems and Control; 2018 SimTech; and 2018 IFAC Symposium on Robust Control Design (2018 RoCond)

4.3 Outreach

- Mario Sznaier served as General Co-Chair of the 2016 IEEE MultiSystems Conference (MSC2016), Program Chair of the 2017 IEEE Conf. Dec. and Control (2017 CDC), and chair of the IFAC Technical Committee on Robust Control (TC 2.5) (2017–present)

Publications

[1] K. Bekiroglu, C. Lagoa, and M. Sznaier. Low-order model identification of mimo systems from noisy and incomplete data. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 4029–4034, Dec 2015.

[2] B. Yilmaz and M. Sznaier. Efficient identification of wiener systems using a combination of atomic norm minimization and interval matrix properties. *54th IEEE CDC*, pages 109 – 114, 2015.

[3] Korkut Bekiroglu, Constantino Lagoa, Stephanie T. Lanza, and Mario Sznaier. System identification algorithm for non-uniformly sampled data. *IFAC-PapersOnLine*, 50(1):7296 – 7301, 2017. 20th IFAC World Congress.

[4] B. Yilmaz, K. Bekiroglu, C. Lagoa, and M. Sznaier. A randomized algorithm for parsimonious model identification. *IEEE Transactions on Automatic Control*, 63(2):532–539, Feb 2018.

[5] O. Camps, C. Lagoa, and M. Sznaier. The role of sparsity and dynamics in extracting information sparsely encoded in very large data sets. In *2016 IEEE Conference on Computational Aspects of Control Systems Design (CACSD)*, pages 398–409, Sept 2016.

[6] Wenqian Liu, Abhishek Sharma, Octavia Camps, and Mario Sznaier. Dyan: A dynamical atoms-based network for video prediction. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[7] O Camps and M Sznaier. The interplay between big data and sparsity in systems identification. In *Geometric and Numerical Foundations of Movements*, pages 133–159. Springer, 2017.

[8] Y. Cheng and M. Sznaier. Identification of lpv systems with lft parametric dependence via convex optimization. *54th IEEE CDC*, pages 1459–1464, 2015.

[9] Rajiv Singh, Mario Sznaier, and Lennart Ljung. A rank minimization formulation for identification of linear parameter varying models. *IFAC-PapersOnLine*, 51(26):74 – 80, 2018. 2nd IFAC Workshop on Linear Parameter Varying Systems LPVS 2018.

[10] Y. Wang, O. Camps, and M. Sznaier. A super-atomic norm minimization approach to identifying sparse dynamical graphical models. *American Control Conference*, 2016.

[11] X. Zhang, O. Camps, and M. Sznaier. Efficient temporal sequence comparison and classification using gram matrix embeddings on a riemannian manifold, 2016 cvpr. *CVPR*, 2016.

[12] X. Zhang, Y. Cheng, Y. Wang, M. Sznaier, and O. Camps. Identification of switched wiener systems based on local embedding. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 3904–3909, Dec 2016.

[13] X. Zhang, M. Sznaier, and O. Camps. Efficient identification of error-in variables switched systems based on riemannian distance-like functions. In *2018 IEEE Conference on Decision and Control (CDC)*, Dec 2018.

[14] Y. Wang, C. Dicle, M. Sznaier, and O. Camps. Self scaled regularized robust regression. In *CVPR*, pages 3261–3269, 2015.

[15] Y. Cheng, Y. Wang, O. Camps, and M. Sznaier. Subspace clustering with priors via sparse quadratically constrained quadratic programming. *CVPR*, 2016.

[16] Mengran Gou, Fei Xiong, Octavia Camps, and Mario Sznaier. Monet: Moments embedding network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[17] Mario Sznaier and Octavia Camps. Sos-rsc: A sum-of-squares polynomial approach to robustifying subspace clustering algorithms. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[18] Y. Cheng, J. Lopez, O. Camps, and M. Sznaier. A convex optimization approach to robust fundamental matrix estimation. In *CVPR*, pages 2170–2178, 2015.

[19] Y. Wang, M. Sznaier, O. Camps, and F. Pait. Identification of a class of generalized autoregressive conditional heteroskedasticity (garch) models with applications to covariance propagation. *54th IEEE CDC*, pages 795–800, 2015.

[20] Yin Wang, Octavia Camps, Mario Sznaier, and Biel Roig Solvas. Jensen bregman logdet divergence optimal filtering in the manifold of positive definite matrices. In *European Conference on Computer Vision*, pages 221–235. Springer, 2016.

[21] X. Zhang, O. Camps, and M. Sznaier. Convex behavioral model (in)validation via jensen-bregman divergence minimization. *American Control Conference*, 2016.

[22] J. A. Lopez, M. Sznaier, and O. Camps. Unsupervised fault detection using semidefinite programming. *54th IEEE CDC*, pages 3798–3803, 2015.

[23] Y. Cheng, M. Sznaier, and C. Lagoa. Robust superstabilizing control design from open loop experimental input/output data. *SySId*, pages 1331–1336, 2015.

[24] Tianyu Dai and Mario Sznaier. Data driven robust superstable control of switched systems. *IFAC-PapersOnLine*, 51(25):402 – 408, 2018. 9th IFAC Symposium on Robust Control Design ROCOND 2018.

[25] T. Dai and M. Sznaier. A moments based approach to designing mimo data driven controllers for switched systems. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 5652–5657, Dec 2018.

[26] Y. Wang, J. Lopez, and M. Sznaier. A convex optimization approach to synthesizing sparse dynamic output feedback controllers. In *IFAC Symposium on Robust Control Design (ROCOND)*, 2015.

[27] Y. Wang, B. Ozbay, M. Sharif, and M. Sznaier. A convex optimization approach to design of information structured linear constrained controllers. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5335–5341, Dec 2016.

[28] Y. Wang, J. Lopez, and M. Sznaier. Convex optimization approaches to information structured decentralized control. *IEEE Transactions on Automatic Control*, 10(63):3393–3403, 2018.

[29] K. Bekiroglu, M. Ayazoglu, C. Lagoa, and M. Sznaier. An efficient approach to the radar ghost elimination problem. In *2016 American Control Conference (ACC)*, pages 3515–3520, July 2016.

[30] K. Bekiroglu, M. Ayazoglu, C. Lagoa, and M. Sznaier. Hankel matrix rank as indicator of ghost in bearing-only tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 54(6):2713–2723, Dec 2018.

[31] C. Dicle, O. Camps, and M. Sznaier. Solving temporal puzzles. *CVPR*, 2016.

[32] R. L. Christensen, K. K. Srensen, R. Wisniewski, and M. Sznaier. Unsupervised fault detection of reefer containers: A moments-based sdp approach. In *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pages 368–373, Aug 2018.

[33] Rasmus L. Christensen, Mario Sznaier, Rafal Wisniewski, and Kresten K. Sorensen. Unsupervised fault detection of refrigeration containers using a mahalanobis inverse moment matrix polynomial. *IFAC-PapersOnLine*, 51(24):249 – 254, 2018. 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2018.

[34] W. Lucia, G. Franz, and M. Sznaier. A hybrid command governor scheme for rotary wings unmanned aerial vehicles. *IEEE Transactions on Control Systems Technology*, pages 1–15, 2018.