2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Worst-Case Optimal Data-Driven Estimators for Switched Discrete-Time Linear Systems

T. Dai M. Sznaier

Abstract— **This paper proposes a data-driven framework to address the worst-case estimation problem for switched discrete-time linear systems based solely on the measured data (input & output) and an ℓ_∞ bound over the noise. We start with the problem of designing a worst-case optimal estimator for a single system and show that this problem can be recast as a rank minimization problem and efficiently solved using standard relaxations of rank. Then we extend these results to the switched case. Our main result shows that, when the mode variable is known, the problem can be solved proceeding in a similar manner. To address the case where the mode variable is unmeasurable, we impose the hybrid decoupling constraint(HDC) in order to reformulate the original problem as a polynomial optimization which can be reduced to a tractable convex optimization using moments-based techniques.**

I. INTRODUCTION

The problem of estimating the value of the output of a system based on previous noisy measurements and a priori information that includes knowledge of both noise and system models is ubiquitous in fields ranging from robotics and navigation to control and video/data analytics, to name just a few. For the case of a single, known system, several solutions have been proposed (depending on the characterization of the dynamics and the noise), including the well-known Kalman filter (KF) [1], extended / unscented Kalman filter [2] [3], particle filter (PF) [4], moving horizon estimation (MHE) [5] and so on. On the other hand, (optimal) estimation for switched systems is considerably less developed.

Roughly speaking, there are two scenarios arising in the context of estimation problems for switched system. The first one assumes full knowledge of the dynamics of the system. In this case, several extensions of the non-switching methods have been proposed to solve the estimation problem. The switching KF [6] assumes Gaussian noise and finds the maximum likelihood estimate using the expectation-maximization (EM) algorithm. The switching PF [7] has the flexibility of choosing the type of noise and models the probability density function using a set of discrete points. [8] proposes the switching MHE by minimizing a receding-horizon quadratic cost function which depends on the recent measurements. In the contex of ℓ_∞ optimal filters, [9] proposed an approach based on the concept of superstability, assuming that the switching sequence is available to the filter. Finally, a worst-case ℓ_∞ optimal filter for the case of unknown switching sequences was proposed in [10], based on polynomial optimization.

This work was supported in part by NSF grants CNS–1646121, CMMI–1638234, IIS–1814631 and ECCS–1808381; AFOSR grant FA9550-19-1-0005; and the Alert DHS Center of Excellence under Award Number 2013-ST-061-ED0001. The authors are with the ECE Department, Northeastern University, Boston, MA 02115. emails: dai.ti@husky.neu.edu, msznaier@coe.neu.edu

The second scenario involves situations where models of the underlying dynamics are not available a priori (an example of this situation is 3D printing, where the underlying processes are typically complex and subject to large parameter variability). In principle this scenario can be handled through a two-step procedure, where a model is identified first, followed by estimation using any of the methods introduced before. However, a potential pitfall of this approach arises from the fact that the identification of switched systems is known to be NP-hard [11]. Typically this forces the use of relaxations that can lead to inaccurate plant models and potentially large errors when using these models for estimation.

To avoid this difficulty, in this paper we propose a data-driven method that produces (worst-case) ℓ_∞ optimal estimators directly from the observed data and some priors, (bounds on the order and number of subsystems and on the noise). In this context, we will consider both the case where the mode variable (e.g. the variable indicating which subsystem is active) is or is not available to the estimator. Motivated by the earlier work [12] which considered the case of a single system and by the ideas in [13], given the experimental data in both cases, we will characterize all the systems consistent with these observations and the given priors in terms of the null space of a matrix constructed from the data. This allows for recasting the optimal switched estimation problem into a rank-constrained optimization which in turn can be relaxed to a convex semi-definite program using the well known nuclear norm surrogate for rank.

The rest of paper is organized as follows: section II introduces the required notations and background knowledge used in the remainder of the paper and formally states the problem of interest. Section III gives the main results of this paper where three cases are discussed: a single system, and switched systems with or without a priori knowledge of the switching sequence. For space reasons, this discussion is confined to the single input single output (SISO) case. Section IV illustrates our results with several academical examples. Finally, section V presents some conclusions and directions for further research.

II. PRELIMINARIES

A. Notation

$\mathbb{R}, \mathbb{N}$	set of real number, set of non-negative integers
$\mathbf{x}, \mathbf{X}$	a vector in $\mathbb{R}^n$, a matrix in $\mathbb{R}^{m \times n}$
$\mathbf{X} \succeq 0$	$\mathbf{X}$ is positive semi-definite
$\|\mathbf{x}\|_\infty$	ℓ^∞-norm of the vector $\mathbf{x} \in \mathbb{R}^n$ $$\|\mathbf{x}\|_\infty \doteq \sup_i |\mathbf{x}(i)|$$
$\sigma_i(\mathbf{X})$	the $\boldsymbol{i}^{th}$ largest singular value of $\mathbf{X}$
diag(X)	create a block diagonal matrix from the matrix $\mathbf{X}$
$\mathcal{E}_\mu(\mathbf{x})$	Expected value of $\mathbf{x}$ with respect to the Borel measure μ
$\boldsymbol{v}_s(\mathbf{x})$	Veronese map of degree s from $\mathbb{R}^n$ to $\mathbb{R}^m$, where $m = \begin{pmatrix} s+n-1 \\ s \end{pmatrix}$. Specifically, $\boldsymbol{v}_s([x_1, \dots, x_n]^T) = [\zeta_1^s, \dots, \zeta_m^s]^T$, where $\zeta^s = x_1^{s_1} x_2^{s_2} \dots x_n^{s_n}, \sum s_i = s$.

B. Information Based Complexity

Next, we recall some concepts from information based complexity(IBC) [14]. These concepts will be used in the paper to define worst-case optimal estimators for switched systems. Consider three linear normed spaces: element, measurement and solution spaces, denoted by Λ, Y and X respectively. We define three linear operators connecting these spaces:

$$\begin{aligned} F: &\quad \Lambda \to Y \quad \text{information operator} \\ S: &\quad \Lambda \to X \quad \text{solution operator} \\ \phi: &\quad Y \to X \quad \text{algorithm} \end{aligned} \tag{1}$$

The goal of estimation is to find an approximation of $S(\lambda) \in X$ for all $\lambda \in \Lambda$. However, only partial information about λ is known, i.e. the measurements of λ given by $F(\lambda)$. In practice, the exact information about $F(\lambda)$ is not available, in this paper, we assume it is perturbed by bounded additive noise $\|\eta\|_\infty \leq \epsilon$. Therefore, the available information is $y = F(\lambda) + \eta \in Y$. The problem of interest is to find an algorithm ϕ giving the closest mapping from measurements y to an estimate of $S(\lambda)$, such that $\|\phi(y) - S(\lambda)\|_\infty$ is minimized. We define the worst-case local error as

$$E(\phi, \epsilon) = \sup_\lambda \|\phi(y) - S(\lambda)\|_\infty \tag{2}$$

The algorithm ϕ_0 is called worst-case locally optimal if $E(\phi_0, \epsilon) = \inf_\phi E(\phi, \epsilon)$, i.e. for every measurement y, the algorithm minimizes the worst-case local error. It is worth noting that $E(\phi_0, \epsilon)$ is also called local radius of information, since it provides the smallest worst case bound of the estimation error among all possible noise.

C. Polynomial Optimization Problems

Polynomial optimization is one of the fundamental problems in the field of optimization and has applications in multiple areas. Given a multivariable polynomial $p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$, with $p_{\boldsymbol{\alpha}}, \boldsymbol{x}^{\boldsymbol{\alpha}}$ denoting the coefficients and monomials of the polynomial, the goal is to minimize $p(\boldsymbol{x})$ subject to a set of polynomial inequalities:

$$\begin{aligned} &p^* = \min_{\boldsymbol{x} \in \mathcal{K}} p(\boldsymbol{x}) \\ &\mathcal{K} = \{\boldsymbol{x} \in \mathbb{R}^n : g_k(\boldsymbol{x}) \geq 0, k = 1, \cdots, N\} \end{aligned} \tag{3}$$

As shown in [15], the above problem is equivalent to minimizing the expectation of the polynomial $p(\boldsymbol{x})$ over all Borel measures μ supported on the semi-algebraic set $\mathcal{K}$, e.g:

$$p^* = \min_{\mu \in \mathcal{P}(\mathcal{K})} \int p(\boldsymbol{x}) \mu(d\boldsymbol{x}) \tag{4}$$

The problem above is an (infinite dimensional) linear program in μ. To obtain finite dimensional relaxations, start by rewriting the equation (4) as:

$$\begin{aligned} &p^* = \min_{\boldsymbol{m}_{\boldsymbol{\alpha}}} \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \boldsymbol{m}_{\boldsymbol{\alpha}} \\ &\mathbf{M} \succeq 0 \\ &\mathbf{L} \succeq 0 \\ &\text{where } \boldsymbol{m}_{\boldsymbol{\alpha}} := \int \boldsymbol{x}^{\boldsymbol{\alpha}} \mu(d\boldsymbol{x}) \end{aligned} \tag{5}$$

Here $\boldsymbol{m}_{\boldsymbol{\alpha}}$ is the $\boldsymbol{\alpha}^{th}$ moment with respect to the Borel measure μ, $\mathbf{M}$ is the moment matrix with entries $\boldsymbol{m}_{\boldsymbol{\alpha}}$ and $\mathbf{L}$ is the localizing matrix with entries $\mathcal{E}_\mu[g_k(\boldsymbol{x})\boldsymbol{x}_{\boldsymbol{\alpha}}]$. Finite dimensional relaxations can be obtained by simply considering truncated versions of the matrices $\mathbf{M}$ and $\mathbf{L}$ containing moments of order up to $2d$, leading to the approximation

$$\begin{aligned} &p_d^* = \min_{\boldsymbol{m}_{\boldsymbol{\alpha}}} \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \boldsymbol{m}_{\boldsymbol{\alpha}} \\ &\mathbf{M}_d \succeq 0 \\ &\mathbf{L}_d \succeq 0 \end{aligned} \tag{6}$$

As shown in [15], $p_d^* \uparrow p^*$ under mild conditions. Moreover, if for some d the so called flat extension condition

$$\text{rank}[\mathbf{M}_d(\mathbf{m})] = \text{rank}\left[\mathbf{M}_{d - \max(\deg(g_k(\boldsymbol{x})))}\right] \tag{7}$$

holds then $p_{\mathbf{m}}^d = p^*$. Specially, when d = 1, it is easy to see that $\text{rank}(\mathbf{M}_1) = 1$ is a sufficient condition for $p_{\mathbf{m}}^d = p^*$.

D. Reweighed Heuristic for Rank Minimization

Given a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, as shown in [16] a convex relaxation of the problem $\min \text{rank}(\mathbf{X})$ is given by the following iterative procedure:

$$\begin{aligned} \min_{\mathbf{Y}, \mathbf{Z}} &\quad \text{Trace}(\mathbf{W_k} \mathbf{H}_{0,k}) \\ \text{subject to} &\quad \mathbf{H}_k \succeq 0 \end{aligned} \tag{8}$$

where

$$\begin{aligned} &\mathbf{H}_{0,k} = \begin{bmatrix} \mathbf{Y}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_k \end{bmatrix} \\ &\mathbf{H}_k = \begin{bmatrix} \mathbf{Y}_k & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Z}_k \end{bmatrix} \\ &\mathbf{W}_k = (\mathbf{H}_{0,k-1} + \delta \mathbf{I})^{-1}, \ \mathbf{W}_1 = \mathbf{I} \end{aligned} \tag{9}$$

E. Statement of Problem

The general problem addressed in this paper is: given (i) experimental measurements (inputs & outputs) corrupted by unknown but bounded noise, and (ii) a priori information consisting of bounds on the number and order of the subsystems and on the norm of the noise, obtain a worst-case optimal estimation of the last output. Formally:

Problem 1: Given a switched output error (SOE) model:

$$\begin{aligned} y_k &= \sum_{i=1}^{n_a} a_i(\gamma_k) y_{k-i} + \sum_{i=1}^{n_b} b_i(\gamma_k) u_{k-i} \\ \tilde{y}_k &= y_k - \eta_k, \quad \|\eta_k\|_\infty \le \epsilon \end{aligned} \tag{10}$$

where the system parameters $a_i(.), b_i(.)$ are unknown, find a worst-case optimal estimate $\hat{y}_k$, of y_k, based on all available data $\tilde{\mathbf{y}} \doteq [\tilde{y}_1, \ldots, \tilde{y}_k]^T$ and $\mathbf{u} \doteq [u_1, \ldots, u_k]^T$. Here u_k, y_k and $\tilde{y}_k$ denote the input, output and its measured value, corrupted by bounded noise η_k, respectively, $\gamma_k = 1, \ldots, s$ is the mode of the system, and n_a, n_b its order.

III. Main Results

In this section, we show that *Problem 1* above can be reduced to a convex optimization. To this effect, we will consider first the case where the system mode $\boldsymbol{\gamma}$ is known. The case of unknown $\boldsymbol{\gamma}$ will be addressed in section III-C.

Define the consistency set $\mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k)$ as the set of all y_k compatible with the existing observations and priors, that is:

$$\begin{aligned} \mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k) \doteq \{ & y_k : \text{there exists some sequence } \boldsymbol{\eta}, \\ & \|\boldsymbol{\eta}\|_\infty \le \epsilon, \text{ and system parameters } a_i(\gamma_k), b_i(\gamma_k) \\ & \text{such that (10) holds}\} \end{aligned} \tag{11}$$

The following result provides the theoretical foundations for the estimators proposed in this paper:

Lemma 1: Define

$$\begin{aligned} y_k^+ &= \max_{\boldsymbol{\eta}, a_i, b_i} y_k \text{ subject to } y_k \in \mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k) \\ y_k^- &= \min_{\boldsymbol{\eta}, a_i, b_i} y_k \text{ subject to } y_k \in \mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k) \\ \hat{y}_k &= \tfrac{1}{2}\left(y_k^+ + y_k^- \right), \end{aligned} \tag{12}$$

then the vector $\hat{y}_k$ is a worst case ℓ_∞ optimal estimator of y_k.

Proof: Follows from Theorem 2.4 in [17] by noting that $\hat{y}_k$ is the Chebyshev center, in the ℓ_∞ sense, of $\mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k)$. ■

A. Estimation for a Single System

To illustrate the main idea of the proposed estimator in a simple setting, we begin by considering the case of a single system. In this case, (10) holds with $\gamma_k = 1$. For ease of notation, define

$$\begin{aligned} \mathbf{c} &= [a_{n_a}, \ldots, a_1, -1, b_{n_b}, \ldots, b_1]^T \\ \mathbf{d}_k &= [y_{k-n_a}, \ldots, y_k, u_{k-n_b}, \ldots, u_{k-1}]^T \end{aligned} \tag{13}$$

It is easy to see that $\mathbf{d}_k^T \mathbf{c} = 0$ holds for all time instants $k > n_a$. Thus, the observed data was generated by a model of the form (10) if and only if the Hankel matrix

$$\mathbf{H} = \begin{bmatrix} \tilde{y}_1 + \eta_1 & \ldots & \tilde{y}_{1+n_a} + \eta_{1+n_a} & u_1 & \ldots & u_{n_b} \\ \tilde{y}_2 + \eta_2 & \ldots & \tilde{y}_{2+n_a} + \eta_{2+n_a} & u_2 & \ldots & u_{1+n_b} \\ \vdots & & \vdots & \vdots & & \vdots \\ \tilde{y}_{k-n_a} + \eta_{k-n_a} & \ldots & \tilde{y}_k + \eta_k & u_{k-n_b} & \ldots & u_{k-1} \end{bmatrix} \tag{14}$$

is rank deficient. It follows that in this case, the consistency set can be rewritten as

$$\begin{aligned} \mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k) \doteq \{ & y_k : \text{there exist some sequence } \boldsymbol{\eta}, \\ & \|\boldsymbol{\eta}\|_\infty \le \epsilon, \text{ such that } \mathbf{H} \text{ is rank deficient}\} \end{aligned} \tag{15}$$

The main advantage of the reformulation above over (11) is that (15) does not contain the unknown parameters a_i, b_i, leading to the following result:

Lemma 2: For the case of a single system, *Problem 1* is equivalent to the following optimization:

$$\begin{aligned} & \eta_{max} = argmax\ \eta_k \quad \text{and } \eta_{min} = argmin\ \eta_k \\ & \text{subject to } \operatorname{rank}(\mathbf{H}) \le R \doteq n_a + n_b \text{ and } \|\boldsymbol{\eta}\|_\infty \le \epsilon \end{aligned} \tag{16}$$

The corresponding worst-case optimal estimator is given by $\hat{y}_k = \tilde{y}_k + \frac{1}{2}(\eta_{max} + \eta_{min})$.

Proof: Follows immediately from the discussion above and the fact that, since $\tilde{y}_k$ is known, maximizing/minimizing y_k is equivalent to maximizing/minimizing η_k ■

Note that the optimization above is non-convex, due to the rank constraint. A convex relaxation can be obtained by combining a line search over η_k with rank minimization, as outlined in Algorithm 1. The main idea here is to choose a value for η_k, minimize the rank of the corresponding $\mathbf{H}$ by using a re-weighted nuclear norm heuristic and increase/decrease the value of η_k until $\mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k)\} \neq \emptyset$.

Algorithm 1 Worst-case Estimation for Single System

```
Initialize: ϵ, δ, step_size(ss), max_iter(mi), W = I
for λ = −ϵ : ss : ϵ do
    for iter = 0 : mi do
        min_η   Trace(W M_0)
        subject to:
            ‖η_{1→k−1}‖_∞ ≤ ϵ
            η_k = λ
            M ⪰ 0
        if σ_{2R+1}(M_0) ≥ δ then
            W = (M_0 + σ_{2R+1}(M_0) I)^{−1}
        else
            return λ
        end if
    end for
end for
where
```

$$\mathbf{M}_0 = \begin{bmatrix} \mathbf{Y} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix}, \qquad \mathbf{M} = \begin{bmatrix} \mathbf{Y} & \mathbf{H} \\ \mathbf{H}^T & \mathbf{Z} \end{bmatrix}$$

$\eta_{min} = \lambda$, to obtain η_{max}, do the line search from the other direction, i.e. $\lambda = \epsilon : -ss : -\epsilon$.

Remark 1: A computationally efficient implementation of the algorithm above can be obtained by considering a receding horizon version where only the previous q measurements are used to predict the next output. Thus, in general this approach may lead to conservative estimates since part of the information is discarded after computing these bounds. However, if the window size is chosen properly, this receding horizon version leads to similar result as Algorithm 1, with a substantial reduction in the computational burden.

B. Estimation for Switched System with Known Switches

Next, we consider switched models of the form (10) in scenarios where the mode variable $\boldsymbol{\gamma}$ can be measured. Proceeding as in the single system case, collect the input/output data in a Hankel matrix $\mathbf{H}$ and define the submatrices $\mathbf{H}_i$, $i = 1, 2, \ldots, s$ where $\mathbf{H}_i$ contains the rows of $\mathbf{H}$ corresponding to time instants where the i^{th} subsystem was active (e.g. those time instants where $\gamma_k = i$). As before, there exists a model of the form (10) that explains the data if and only if there exist an admissible noise sequence such that all the matrices $\mathbf{H}_i$ are rank deficient. Equivalently, in this case the consistency set is given by:

$$\mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k) \doteq \{y_k : \text{there exist some sequence } \boldsymbol{\eta}, \\ \|\boldsymbol{\eta}\|_\infty \leq \epsilon, \text{ such that } \mathbf{H}_i \text{ are rank deficient}, i = 1, \ldots, s\} \tag{17}$$

Thus, an optimal estimator can be found by solving:

$$\begin{aligned} &\eta_{max} = argmax\ \eta_k \quad \text{and} \quad \eta_{min} = argmin\ \eta_k \\ &\text{subject to rank } (\mathbf{D}) \leq R \doteq \sum_{\gamma_k=1}^{s} n_a(\gamma_k) + n_b(\gamma_k) \\ &\text{and } \|\boldsymbol{\eta}\|_\infty \leq \epsilon, \text{ where } \mathbf{D} \doteq \mathbf{diag}(\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_s) \end{aligned} \tag{18}$$

Problem (18) can be efficiently solved using an algorithm similar to Algorithm 1, with $\mathbf{H}$ and R replaced by $\mathbf{D}, R$ as defined above.

C. Estimation for Switched System with Unknown Switches

When the mode variable γ is unknown, the problem becomes non-trivial. In this case, since the data segmentation is not available, one cannot proceed as in section III-B to build the submatrices $\mathbf{H}_i$ and impose rank deficiency of $\mathbf{D}$. However, in this case, as noted in [18], since the data point always belongs to one of the subsystem, the so called hybrid decoupling constraint (HDC) always holds, i.e.

$$\prod_{\gamma=1}^{s} \mathbf{d}_k^T \mathbf{c}_\gamma = 0 \tag{19}$$

where

$$\mathbf{c}_\gamma = [a_{n_a}(\gamma), \ldots, a_1(\gamma), -1, b_{n_b}(\gamma), \ldots, b_1(\gamma)]^T \tag{20}$$

Note that, for each k, the equation above can be written in terms of an homogeneous multivariate polynomial of degree s in the variables $\mathbf{d}_k$ of the form

$$\boldsymbol{v}_s^T(\mathbf{d}_k)\hat{\mathbf{c}}_\gamma = 0$$

where $\boldsymbol{v}_s(.)$ denotes the Veronese map of degree s, and $\hat{\mathbf{c}}_\gamma$ is a vector related to $\mathbf{c}_\gamma$. Thus, given the experimental data, there exist some switching sequence $\boldsymbol{\gamma}$ such that a model of the form (10) explains the data if and only if the matrix $\mathbf{V}_s$ formed by stacking the vectors $\boldsymbol{v}_s^T(\mathbf{d}_k)$ is rank deficient. To illustrate the idea, consider the following example

Example 1: Assume $(n_a, n_b) = (1, 1)$, s = 2, then $\mathbf{c}_\gamma = [a_1(\gamma), b_1(\gamma)]^T (\gamma = 1, 2)$, $\mathbf{d}_k = [y_{k-1}, y_k, u_{k-1}]^T$. In this case, the corresponding Veronese map $\boldsymbol{v}_s(\mathbf{d}_k)^T$ and $\hat{\mathbf{c}}_\gamma$ are given by

$$\boldsymbol{v}_s(\mathbf{d}_k)^T = [y_k^2, \quad -y_k y_{k-1}, \quad -y_k u_{k-1}, \\ y_{k-1}^2, \quad u_{k-1}^2, \quad y_{k-1}u_{k-1}] \tag{21}$$

$$\hat{\mathbf{c}}_\gamma = [1, a_1(1) + a_1(2), b_1(1) + b_1(2), a_1(1)a_1(2), \\ b_1(1)b_1(2), a_1(1)b_1(2) + a_1(2)b_1(1)]^T \tag{22}$$

Lemma 3: For the case of switched systems with an unknown switching sequence, *Problem 1* is equivalent to the following optimization:

$$\begin{aligned} &\eta_{max} = argmax\ \eta_k \quad \text{and} \quad \eta_{min} = argmin\ \eta_k \\ &\text{subject to rank}(\mathbf{V}_s) \leq R \text{ and } \|\boldsymbol{\eta}\|_\infty \leq \epsilon \\ &\text{where} \quad R \doteq \binom{n_a + n_b + s}{n_a + n_b} - 1 \end{aligned} \tag{23}$$

The corresponding worst-case optimal estimator is given by $\hat{y}_k = \tilde{y}_k + \frac{1}{2}(\eta_{max} + \eta_{min})$

Proof: Follows from the discussion above by noting that in this case

$$\mathcal{T}(\tilde{\mathbf{y}}, \mathbf{u}, \epsilon, k) \doteq \{y_k : \text{there exist some sequence } \boldsymbol{\eta}, \\ \|\boldsymbol{\eta}\|_\infty \leq \epsilon, \text{ such that } \mathbf{V}_s \text{ is rank deficient}\} \tag{24}$$

■

Note that, as stated, the optimization (23) is intractable, since it requires minimizing the rank of a matrix that depends polynomially in the variables. Further, since rank is not a semi-algebraic function, standard polynomial (or moment) based tools cannot be used. Nevertheless, as we show next, this problem can be relaxed to a tractable convex optimization.

Theorem 1: Let $\mathbf{m}_{\boldsymbol{\alpha}}$ denote the moment sequence corresponding to the noise variable $\boldsymbol{\eta}$ and $\tilde{\mathbf{V}}_s(\mathbf{m}_{\boldsymbol{\alpha}})$ denote the matrix obtained by replacing each $\boldsymbol{\alpha}^{th}$ degree monomial $\boldsymbol{\eta}^{\boldsymbol{\alpha}}$ in $\mathbf{V}_s(\boldsymbol{\eta})$ with the corresponding $\boldsymbol{\alpha}^{th}$ order moment $\mathbf{m}_{\boldsymbol{\alpha}}$. Consider the following rank constrained minimization problem (affine in the variables $\mathbf{m}_{\boldsymbol{\alpha}}$):

$$\begin{aligned} &\tilde{\eta}_{max} = argmax\ \eta_k \quad \text{and} \quad \tilde{\eta}_{min} = argmin\ \eta_k \\ &\text{subject to:} \\ &\text{rank}(\tilde{\mathbf{V}}_\mathbf{s}) \leq R, \ \ |m_t| \leq \epsilon, \ t = 1, \ldots, k-1 \text{ and} \\ &\text{rank}(\mathbf{M}_\eta) = 1 \end{aligned} \tag{25}$$

where m_t denotes the entry of the moment matrix $\mathbf{M}_\eta$ corresponding to $\mathcal{E}(\boldsymbol{\eta})$ Then, $\tilde{\eta}_{max} = \eta_{max}$ and $\tilde{\eta}_{min} = \eta_{min}$.

Proof: (only sketch given due to space constraints) From Theorem 2 in [13], there exist a sequence $\boldsymbol{\eta}$, $|\boldsymbol{\eta}| \leq \epsilon$ that renders $\mathbf{V}$ rank deficient if and only if there exist a sequence of moments $\mathbf{m}$ with a representing measure supported in $[-\epsilon, \epsilon]$ such that $\tilde{\mathbf{V}}_s(\mathbf{m})$ is rank deficient. The rank 1 constraint on $\mathbf{M}_\eta$ and $|m_t| < \epsilon$ in (25) are sufficient conditions for the existence of this measure. ■

Note that (25) is now a constrained rank minimization subject to semidefinite constraints, where all the matrices involved are affine in the variables $\mathbf{m}$. Therefore, as before, a convex relaxation can be obtained by replacing rank with a (weighted) nuclear norm. The resulting problem can be solved using an algorithm similar to Algorithm 1, replacing $\mathbf{H}, R$ by $\mathbf{P} \doteq \textbf{diag}\,(\tilde{\mathbf{V}}_s(\mathbf{m}), \mathbf{M}_\eta)$ and $S = R + 1$.

IV. Experiments

In this section we illustrates the proposed filtering framework with several academic examples. We start from the estimation for a single system. The model we used to generate the data is

$$y_k = 1.1 y_{k-1} - 0.28 y_{k-2} + 0.3 u_{k-1} + 0.2 u_{k-2} \qquad (26)$$

The initial output is $y = [1, 2]$. The input $\|\mathbf{u}\|_\infty \leq 1$ and the noise $\|\boldsymbol{\eta}\|_\infty \leq \epsilon$ are uniformly distributed. We chose the following values for the hyper-parameters: $\epsilon = 0.1, \delta = 1e-4, ss = 0.004, mi = 5$. The results of applying Algorithm 1 are shown in Fig. 1, where the green line segment gives the noise bound.

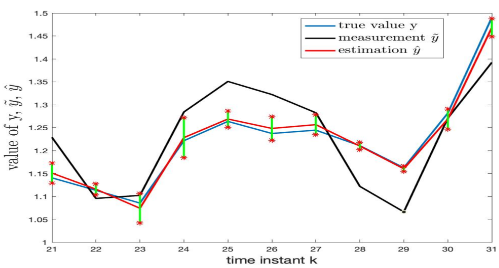

Fig. 1. Estimation for a Single System

The results of using the corresponding receding-horizon version of the algorithm with window size 21 are shown in Fig. 2

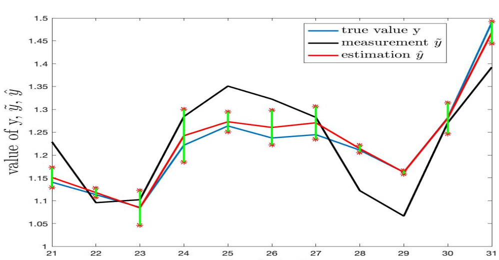

Fig. 2. Receding Horizon Estimation for a Single System (T=21)

As expected the receding-horizon version gives a larger estimation error since it only uses the previous 21 measurements for prediction, and hence is more conservative. Increasing the window size reduces the conservatism at the price of an increased computational cost.

For the switched system with a known switching sequence, we use the following model consisting of two subsystems:

$$\begin{aligned} y_k &= y_{k-1} - 0.24 y_{k-2} + 0.2 u_{k-1} + 0.6 u_{k-2} \\ y_k &= 1.2 y_{k-1} - 0.35 y_{k-2} + 0.4 u_{k-1} + 0.3 u_{k-2} \end{aligned} \qquad (27)$$

In this case the switch sequence was generated randomly without any dwell time assumptions. The estimates obtained solving a convex relaxation of (18) are shown in Fig. 3

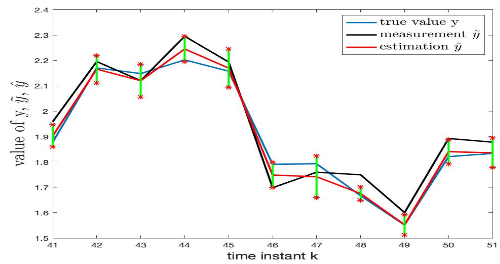

Fig. 3. Estimation for a Switched System (Known Switches)

with the results of the corresponding receding-horizon version (size 41) shown in Fig. 4

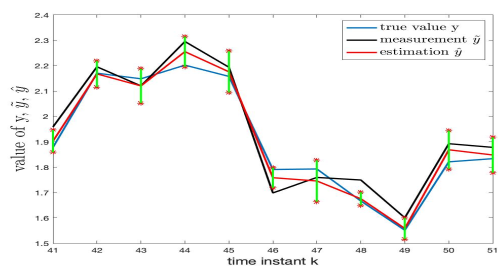

Fig. 4. Receding Horizon Estimation for a Switched System (Known Switches, T=41)

Finally, consider the following switched system with unknown switches:

$$\begin{aligned} y_k &= 0.6 y_{k-1} + 0.4 u_{k-1} \\ y_k &= 0.8 y_{k-1} + 0.5 u_{k-1} \end{aligned} \qquad (28)$$

For this experiment, we pick as initial value $y = 1$ and changed mi to be 50, since rank minimization of a moment matrix is slow to converge. The results obtained by solving a convex relaxation of (25) are shown in Fig. 5

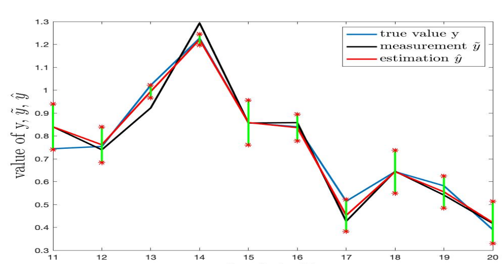

Fig. 5. Estimation for a Switched System (Unknown Switches)

To assess the conservatism introduced by not being able to measure the mode variable, we solved (18) on the same model, using information about which subsystem is active at a given time instant, leading to the results shown in Fig. 6

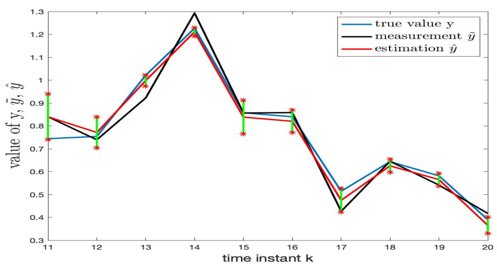

Fig. 6. Estimation for Switched System (Known Switches)

As expected, comparing Figures 6 and 5 shows that exploiting knowledge about the switching sequence leads to a tighter bound of the estimate (and thus a smaller estimation error).

V. Conclusion

This paper proposes a data-driven framework to address the worst-case estimation problem for switching discrete-time linear systems based only on the experimental measurements and some minimal a priori information. We considered three scenarios (single system/switching system with known/unknown switches) and showed that in all cases ℓ_∞ worst-case optimal estimators can be obtained by solving a rank-constrained optimization. Efficient convex relaxations of these problems can be obtained by replacing rank with its nuclear norm surrogate. Further computational complexity can be achieved by considering a receding-horizon version of the algorithm, where information from the past is encapsulated in bounds for the observed variables. As illustrated with examples, this receding horizon version achieves good performance when the size of the sliding window is chosen appropriately. The main advantage of the proposed method is that it avoids a systems identification step (with the potential entailed conservatism), producing worst-case optimal estimates directly from the data. Further, the proposed algorithm can be easily extended to the case of missing data (for instance due to sensor failure) by simply treating this missing data as an unknown in the optimization problem.

A potential difficulty of the proposed algorithm in the case of unknown switches (shared with systems identification algorithms for this case) is the non-trivial computational burden, especially in cases where the number or order of the subsystems is not small. Future research will seek to exploit the structure of the matrix $\mathbf{V}$ (e.g. it can be shown that the aggregate sparsity pattern of the graph associated with this matrix has cliques which have at most size n_a) together with recent results on rank-preserving matrix completions to substantially reduce this complexity. Promising preliminary results seem to indicate that this approach can lead to algorithms that scale linearly with the number of data points and subsystems.

References

[1] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[2] Lennart Ljung. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1):36–50, 1979.

[3] Eric A Wan and Rudolph Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee, 2000.

[4] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188, 2002.

[5] Frank Allgöwer, Thomas A Badgwell, Joe S Qin, James B Rawlings, and Steven J Wright. Nonlinear predictive control and moving horizon estimation—an introductory overview. In *Advances in control*, pages 391–449. Springer, 1999.

[6] Kevin P Murphy. Switching kalman filters. 1998.

[7] Franois Caron, Manuel Davy, Emmanuel Duflos, and Philippe Vanheeghe. Particle filtering for multisensor data fusion with switching observation models: Application to land vehicle positioning. *IEEE transactions on Signal Processing*, 55(6):2703–2719, 2007.

[8] Angelo Alessandri, Marco Baglietto, and Giorgio Battistelli. Receding-horizon estimation for switching discrete-time linear systems. *IEEE Transactions on Automatic Control*, 50(11):1736–1748, 2005.

[9] M. Sznaier, B. Yilmaz, and F. Blanchini. Bounded complexity filters for switched systems. In *Proceedings of the 2011 American Control Conference*, pages 2006–2011, June 2011.

[10] Y. Cheng, Y. Wang, and M. Sznaier. worst-case optimal estimators for switched arx systems. In *52nd IEEE Conference on Decision and Control*, pages 4036–4041, Dec 2013.

[11] Necmiye Ozay, Mario Sznaier, and Constantino Lagoa. Convex certificates for model (in) validation of switched affine systems with unknown switches. *IEEE Transactions on Automatic Control*, 59(11):2921–2932, 2014.

[12] Tao Ding, Mario Sznaier, and Octavia Camps. Receding horizon rank minimization based estimation with applications to visual tracking. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 3446–3451. IEEE, 2008.

[13] Necmiye Ozay, Constantino Lagoa, and Mario Sznaier. Set membership identification of switched linear systems with known number of subsystems. *Automatica*, 51:180–191, 2015.

[14] Joseph F Traub. *Information-based complexity*. John Wiley and Sons Ltd., 2003.

[15] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001.

[16] Maryam Fazel, Haitham Hindi, and Stephen P Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of the 2003 American Control Conference, 2003.*, volume 3, pages 2156–2162. IEEE, 2003.

[17] J. F. Traub and H. Wozniakowski. *A general theory of optimal algorithms*. Academic Press., 1980.

[18] René Vidal, Stefano Soatto, Yi Ma, and Shankar Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 1, pages 167–172. IEEE, 2003.